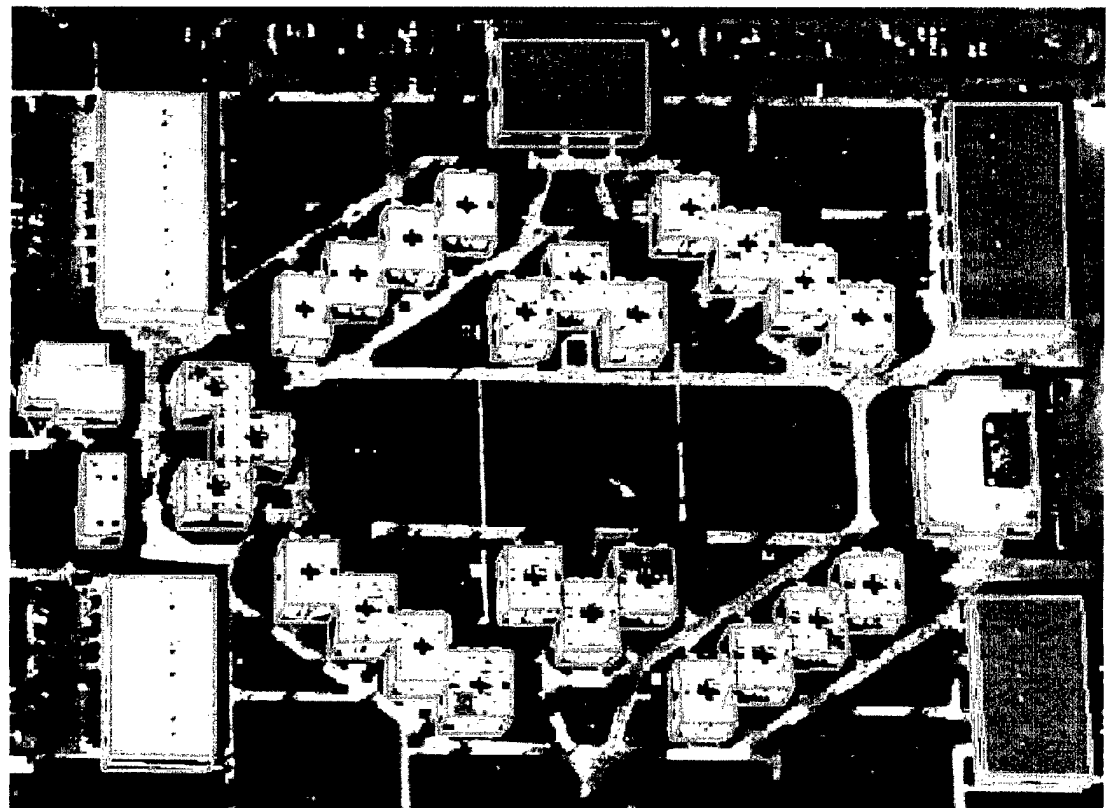US Army Corps
of Engineers®
Engineer Research and
Development Center

Research in Knowledge-Based Automatic Feature Extraction

June 2000

University of Southern California:

R. Nevatia and A. Huertas

20000802 184

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE July 2000	3. REPORT TYPE AND DATES COVERED Final Technical

4. TITLE AND SUBTITLE

Research in Knowledge-Based Automatic Feature Extraction

5. FUNDING NUMBERS

DACA76-97-K-0001

6. AUTHOR(S)

R. NEVATIA and A. HUERTAS

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Southern California
Institute for Robotics and Intelligent Systems
Powell Hall 204
Los Angeles, CA 90089-0273

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

DARPA
3701 No. Fairfax Dr.
Arlington, VA 22203-1714

U.S. Engineer Research Development Center
Topographic Engrg. Center
7701 Telegraphic Rd.
Alexandria, VA 22315-3864

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ERDC/TEC CR-00-2

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report provides a description of USC's activities as part of the DARPA Automated Population of Geospatial Databases (APGD) project. Focus of this effort is on detection, delineation, and 3-D description of buildings from aerial images. Buildings are important in all environments and accurate models of them are needed for tasks such as mission planning and rehearsal, tactical training, damage assessment, and change detection. An overview of a system for modeling buildings using multiple views is given and its performance is quantitatively evaluated. The automatic system is limited to rectilinear-shaped buildings with flat or symmetric gabled roofs, and assumes that ground height, camera models, and illumination parameters are known. Use cues from IFSAR data is proposed. It is shown that such cues can not only greatly improve the efficiency of the automatic system but also improve the quality of the results. The resolution of such data, however, is typically lower than that of electro-optical images and the data may have missing or erroneous elements. An approach that allows a user to assist the automatic system in modeling buildings is described. The approach is designed to be efficient in user time and effort while preserving the quality of the models created. The assisted system also is able to handle multicomponent buildings, buildings having several layers, and buildings having arbitrary shapes. Simple models can be created by only one or two clicks in many cases. Efficient editing of automatically derived models also is possible. Results on images of Ft. Hood, TX, Ft. Benning, GA, and Washington D.C. sites are presented.

14. SUBJECT TERMS

3-D Model construction, building detection, sensor fusion

15. NUMBER OF PAGES

52

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UNLIMITED

TABLE OF CONTENTS

ILLUSTRATIONS

TABLES

Preface

Acknowledgments

RESEARCH IN KNOWLEDGE-BASED AUTOMATIC FEATURE EXTRACTION

1. Introduction and Overview

The goal of this project is to develop feature extraction methods for automated population of geospatial databases (APGD) with particular focus on buildings. The task includes detection, delineation, and description of 3-D buildings. Buildings are objects of obvious importance in urban environments and accurate models of them are needed for a number of battlefield awareness tasks such as mission planning, mission rehearsal, tactical training, and damage assessment. Other applications include intelligence analysis for site monitoring and change detection.

The problem of 3-D feature extraction has many sources of difficulties including those of segmentation, 3-D inference, and shape description. Segmentation of buildings is difficult because of the presence of large numbers of objects such as roads, sidewalks, landscaping, markings, and shadows near the buildings and the presence of texture on building surfaces. Three-dimensional information is not explicit in an intensity image; its inference from multiple images requires finding correct corresponding points or features in two or more images. Direct ranging techniques such as interferometric synthetic aperture radar (IFSAR [Curlander & McDonough, 1991, Jakowatz et al., 1996]) can provide highly useful 3-D data though the data typically have areas of missing elements and may contain some points with grossly erroneous values. Once the objects have been segmented and 3-D shape recovered, the task of shape description still remains. This consists of forming complex shapes from simpler shapes that may be detected at earlier stages. For example, a building may have several wings, possibly of different heights, that may be detected as separate parts rather than one structure initially.

The approach used in this effort is to use a combination of tools: reconstruction and reasoning in 3-D, use of multiple sources of data and perceptual grouping. Context and domain knowledge guide the application of these tools. Context comes from knowledge of camera parameters, geometry of objects to be detected, and illumination conditions (primarily the sun position). Some knowledge of the approximate terrain also is used. The information from sensors of different modalities, such as IFSAR and electro-optical panchromatic (PAN), is fused not at pixel level but at higher feature levels. Our approach also allows for integration of information from multispectral images though this is not being pursued actively as part of the described effort. Other approaches recently reported [Hoepfner et al., 1997] use the IFSAR data to fit models of roof surfaces at regions of interest.

The described system is limited to buildings with rectilinear shapes. Most of our work has been on buildings with flat roofs but the system also can handle buildings with symmetrical slanted roofs (gables). It is assumed that camera models are given and approximated by orthographic projection locally, that the ground is flat with known height, and the sun position is given (computable from latitude, longitude, and time-of-day). Multiple images are *not* assumed to have been taken at the same time. The multiview system is described briefly in Section 1.1. Incorporation of cues from IFSAR is described in Section 1.2 and a comparative system evaluation is given in Section 1.3. A more detailed description of cue analysis is given in Section 2 and in [Huertas et al., 1998].

A user can interact with the described system, either to edit the results of the automatic system or to provide cues for it. Recent additions to this system allow modeling of non-rectilinear or polygonal shapes as well. The aim is to make the user input efficient, requiring much less effort than would be necessary for conventional interactive systems that largely take care only of geometric computations and bookkeeping. The assisted system is described briefly in Section 1.4 with more details given in [Li et al., 1998].

1.1 Multiview System (MVS)

A number of systems that use multiple views have been described in the literature [Jaynes et al., 1997; Collins et al., 1998; Roux & McKeown, 1994]. The system described in this report derives from an earlier system described in [Noronha & Nevatia, 1997]; a block diagram is shown in Figure 1.1. The approach is basically one of hypothesize and verify. *Hypotheses* for potential roofs are made from fragmented lower-level image features. The system is hierarchical and uses evidence from all the views in a non-preferential, order-independent way. Promising hypotheses are *selected* among these by using relatively inexpensive evidence from the rooftops only. The selected hypotheses are then *verified* by using more reliable global evidence. The verified hypotheses are then examined for overlap that may result in either elimination or in the merging of them. Cues from a depth map (such as IFSAR or a DEM) can be incorporated at the hypotheses formation, selection, or verification stages.
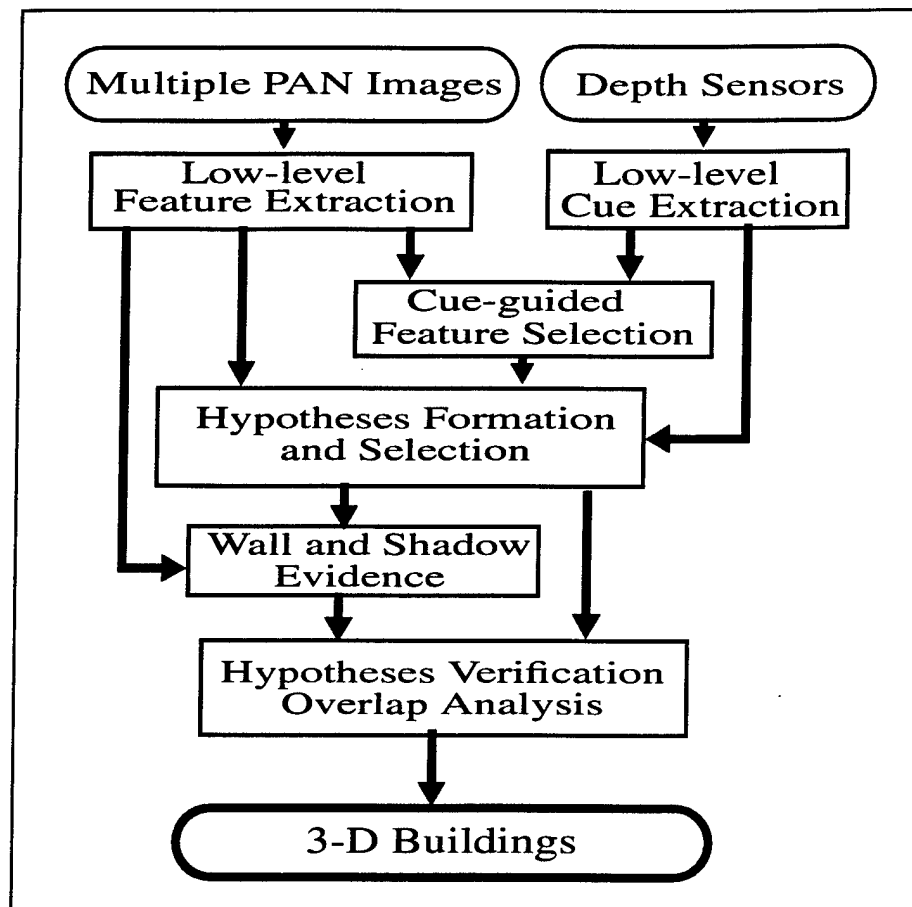
Figure 1.1 Block diagram of the system

This system is designed for rectilinear buildings; complex buildings are decomposed into rectangular parts. Flat rooftops project to parallelograms in the images (the projection is nearly orthographic over the scale of a building), gables project to a pair of parallelograms sharing a side. Lines, junctions, and parallel lines are the basic features used to form roof hypotheses. Flat roof hypotheses are formed by a pair of parallel lines and *U* structures (U's represent three sides of a parallelogram). Gable hypotheses are formed from a *triple* of parallel lines. Closed hypotheses are formed from these features by using the best available image lines if any, or else closures are synthesized from the ends of the parallel lines.

2

Three-dimensional roof hypotheses could be inferred from the 2-D hypotheses by using line matches; however, the line matches are often not necessarily unique. Instead, an estimate for the heights of roof lines is made by conducting a search. For a flat roof only a single height needs to be determined, for the symmetric gables we need to find two heights. For each height estimate, the corresponding 3-D hypotheses is projected in each other view, and line evidence for each projection is computed. The evidence consists of the sum of the lengths of the supporting segments. The heights with the best evidence are selected.

The hypothesis formation process is rather liberal and a large number of hypotheses are typically formed at this stage. A smaller set is *selected* using the underlying image evidence for the roof hypotheses. Positive evidence comes from lines near the projected hypotheses, negative evidence comes from lines crossing the hypotheses. A coarse analysis also is applied to select among overlapping hypotheses. Currently, the selection process is applied to the flat roof cases only.

The next step is to *verify* whether the selected hypotheses have additional evidence for corresponding to representing buildings. This evidence is collected from the roof, the walls, and the shadows that should be cast by the building. Since the hypotheses are represented in 3-D, deriving the projections of the walls and shadows cast, and determining which of these elements are visible from the particular view point is possible. These in turn guide the search procedures that look in the various images for evidence of these elements among the features extracted from the image. A score is computed for each evidence element.

Each of the collected evidence parameters is composed of smaller pieces of evidence. A critical question is how to combine these small pieces of evidence to decide whether a building is present or not and how much confidence should be put in it. A variety of methods for this is available such as linear weighted sums of components, decision trees, certainty theory, neural networks, and statistical classifiers. The results shown in this report use a *decision tree* classifier. We also are investigating use of a Bayesian classification approach in a separately funded project.

After verification, several overlapping verified hypotheses may remain. Only one of the significantly overlapping hypotheses is selected. The overlap analysis procedure examines not only the evidence available for alternatives but, also separately, the evidence for components that are not common.

The system currently detects flat and gabled roofed buildings separately. Figures 1.2 and 1.3, respectively, show the flat roofed and gabled roofs detected for the McKenna MOUT site at Ft. Benning, GA, by using two stereo images.

Figure 1.4 shows the combination of these two results after applying overlap analysis to eliminate conflicts. The results show that all the buildings are detected, at least in part, though not all are completely accurate. Some portions of gable roofs are detected as flat and vice-versa. The combination process that analyzes the overlap between these is in a preliminary stage of development. The roof on the lower right is detected as both a gable and a flat roof but the flat hypothesis dominates, incorrectly, in this case. The gabled portions on the upper center are not detected as gabled but as flat roofs. There are two small flat "false alarms" (*i.e.*, buildings found where there are none) adjacent to the flat building in the center of the right part of the image. Another false alarm is present between this and the adjacent building. These are due to the walls that extend beyond the building sides.

1.2 Cues From IFSAR

The performance of the building detection and description system can be greatly improved if a source of direct range information is available. IFSAR data have become available in recent years. In addition to reflectivity information, it also contains information of 3-D points in a scene.

The resolution of the IFSAR images is more limited and many wrong values are present due to the reflective properties of the surface material in the radar spectrum; it is preferable to use IFSAR for detection and panchromatic images for accurate delineation. Cues from IFSAR data can be used in a number of ways: in se-
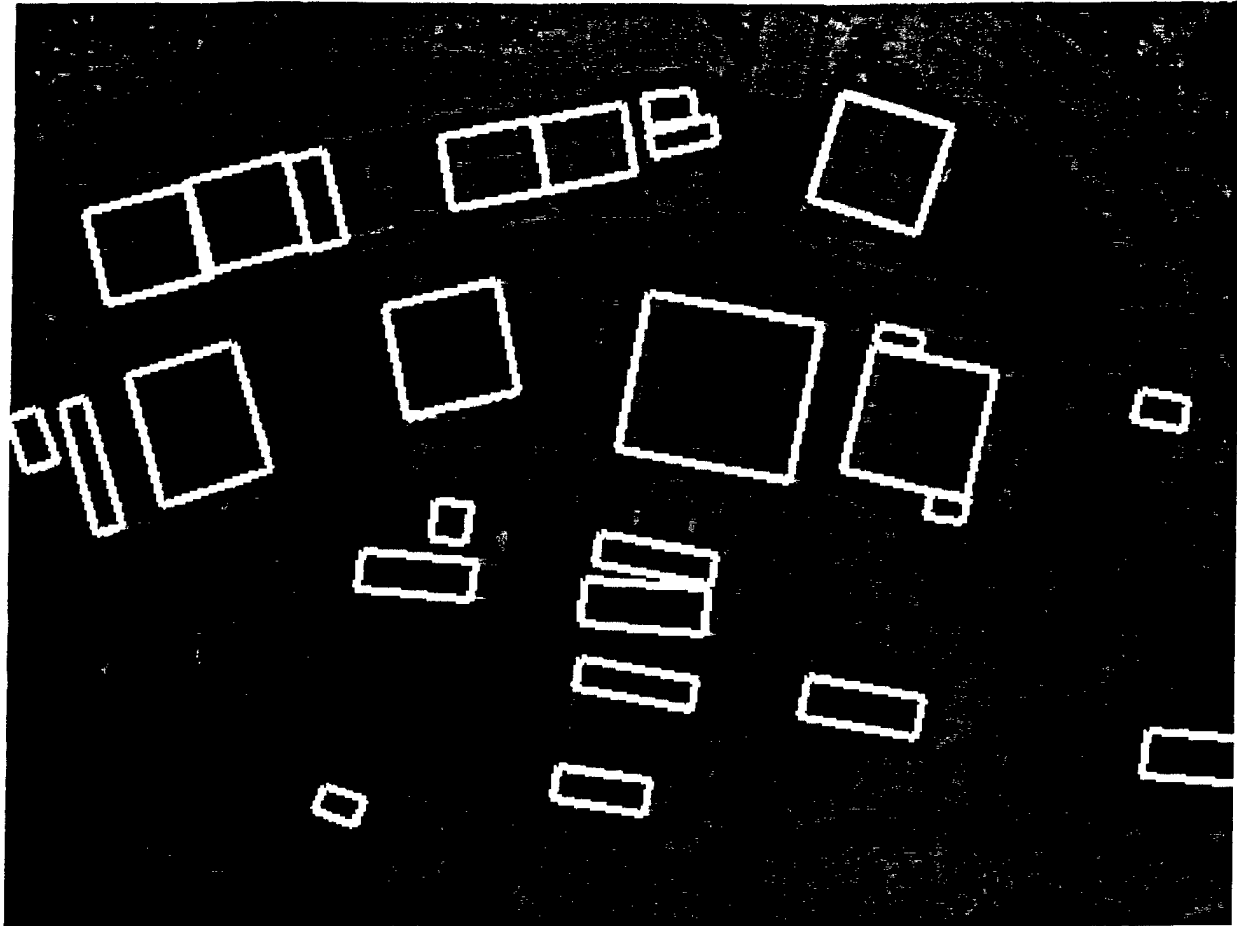
Figure 1.2 Flat-roofed building components extracted using PAN images only

lecting areas to process where buildings may be present, in eliminating certain building hypotheses, and in adding confidence to the presence of buildings.

IFSAR data are given in the form of three images, called the **mag, dte,** and **cor** images corresponding to the reflected magnitude, digital terrain elevation, and phase correlation respectively. For certain kinds of sensors, such as a searchlight mode Scandia sensor, cues for buildings can be derived from the **dte** data alone. A **dte** image for the Ft. Benning site is shown in Figure 1.5. Regions that may correspond to buildings, shown in Figure 1.6, are derived by convolving the image with a Laplacian-of-Gaussian filter that smooths the image and locates the object boundaries by the positive-valued regions bounded by the zero-crossings in the convolution output [Huertas et al., 1998].

The nearby trees on the south and west sides of the site also are well represented. For other kinds of sensors such as IFSARE, for which we have data over the Ft. Hood site, we have found it useful to use all three images (**mag, dte** and **cor**) for extracting the cues [Huertas et al., 1998].

Object cues are used in several ways and at different stages of the hypotheses formation and validation processes. Figure 1.7 shows the linear structures that are near the cue regions. By using lines near objects, the system not only is more efficient as it processes a smaller number of features, but these, presumably, the more relevant features, lead to better hypotheses. We also use these cues to help select promising hypotheses, or conversely, to help disregard hypotheses that may not correspond to objects.

4
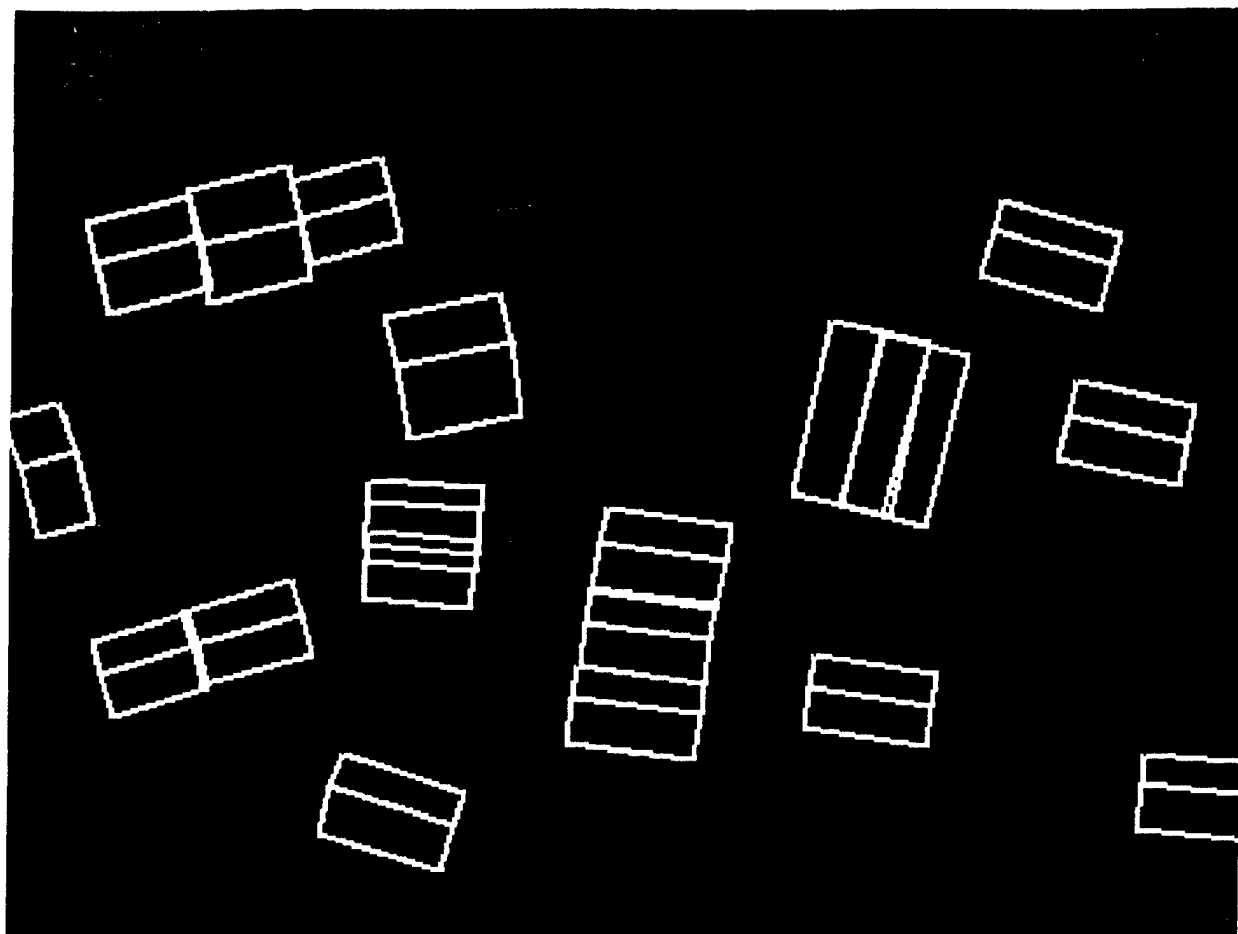
Figure 1.3 Gabled-roofed building components extracted using PAN images only
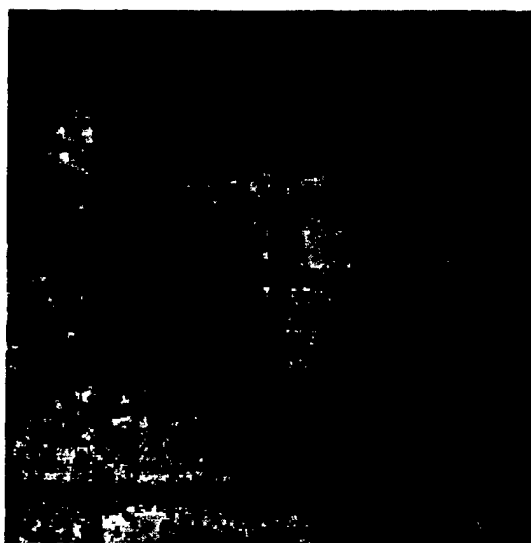
Figure 1.5 IFSAR derived DEM image

5
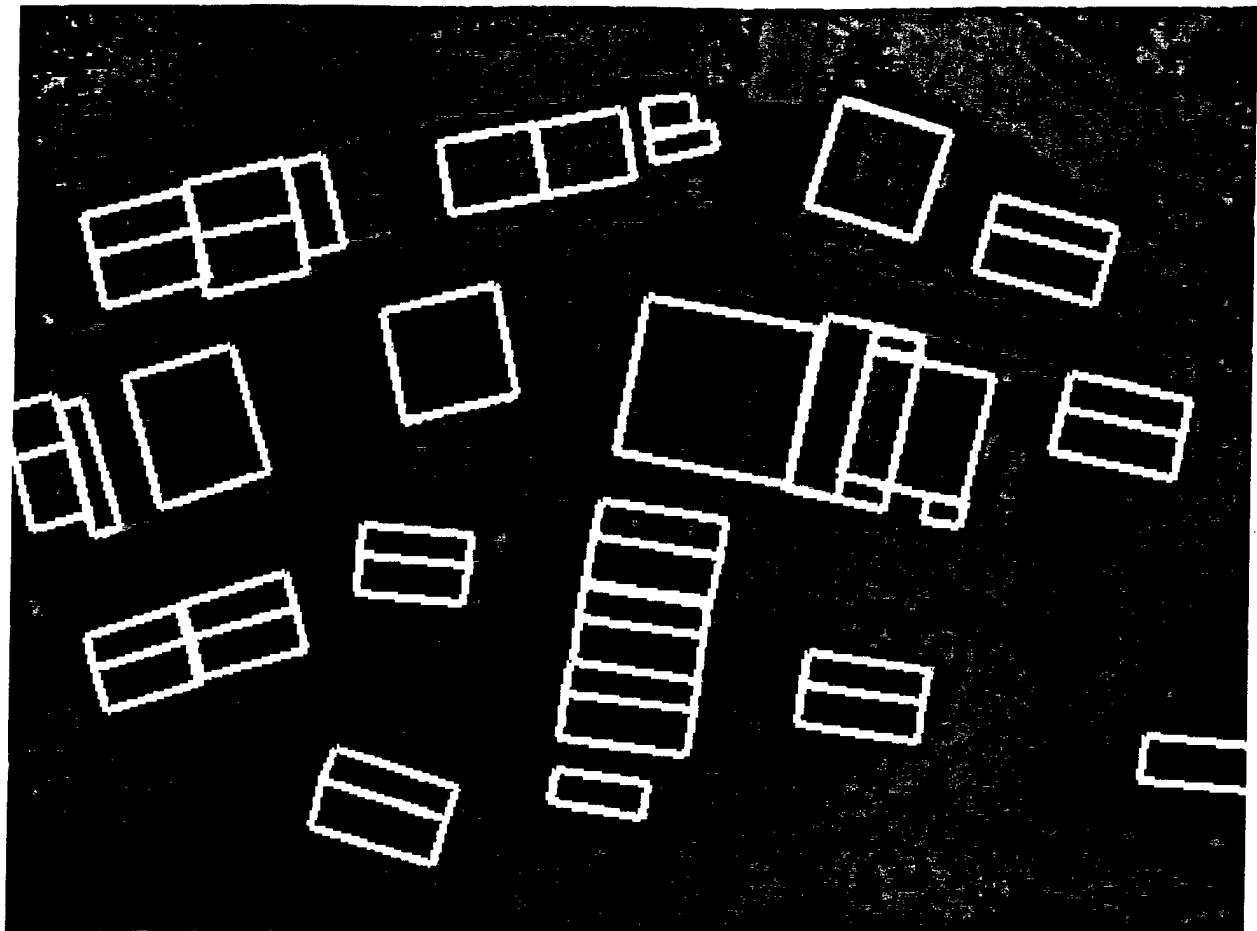
Figure 1.4 Final automatic result for MOUT site, using PAN images only

Just as poor hypotheses can be discarded because they lack IFSAR support, the ones that have a large support see their confidence increase during the verification stage.

Figures 1.8 and 1.9 show the detected flat and gabled-roofed buildings using the IFSAR cues.

Figure 1.10 shows the combined flat and gable verified hypotheses. This result shows no false alarms. Also, the roofs of the gabled buildings are detected correctly; however, parts of the gabled buildings in the upper center have not been detected.

1.3 System Evaluation

Quantitative evaluation of system performance is important in determining its utility. We define some metrics below that are similar to those contained in various proposals though there is not yet a complete agreement on the most desirable ones [McKeown et al., 1997; Fischler et al., 1998]. All comparisons are made with a reference model that may be derived by a human operator or by measurements on the ground. The following terms are defined:

- *TP* (True Positive): a detected feature that also is in the reference.

- *FP* (False Positive): a detected feature that is not in the reference; also termed a *false alarm*.

- *FN* (False Negative), a feature in the reference that is not detected.

Figure 1.6 Cues extracted from IFSAR DEM

Figure 1.7   Lines near IFSAR cues
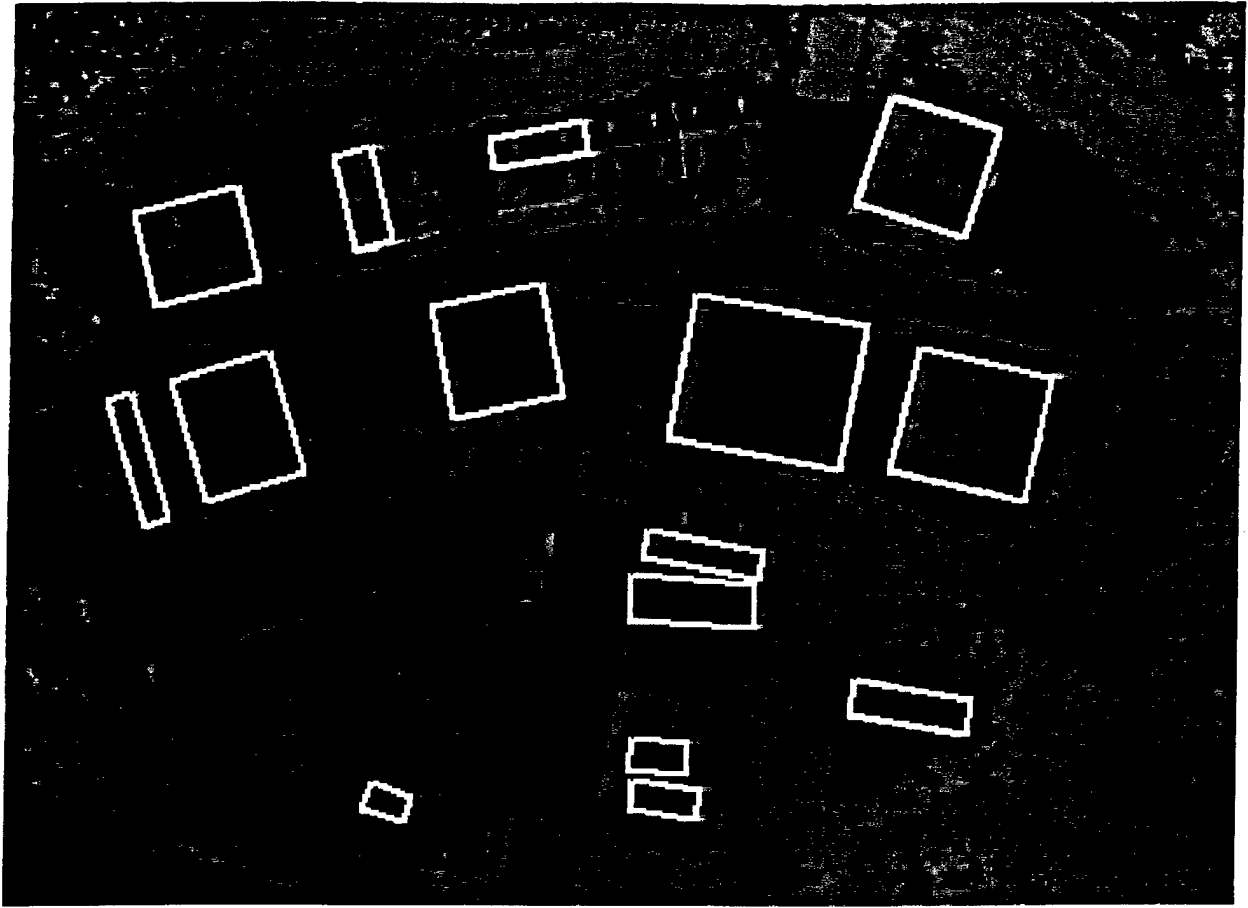
Figure 1.8 Flat-roofed building components extracted using IFSAR cueing

These quantities are combined to give:

$$Detection\ Rate\ =\ \frac{TP}{(TP + FN)}$$

$$False\ Alarm\ Rate\ =\ \frac{FP}{(TP + FP)}$$

Note that with these definitions, the detection rate is computed as a fraction of the reference features whereas the false alarm rate is computed as a fraction of the detected features.

In the definitions given above, a feature could be an object, an area element, or a volume element. The disadvantage of using image pixels (*i.e.*, the roof area) is that the numbers from a few large buildings may dominate a number of smaller buildings. If they are to be computed on the basis of an entire object, then we need to define when we consider an object to have been detected. In the evaluations, a building is considered to have been detected if *any* part of it has been detected. The amount by which a building has been correctly detected is computed by the number of points inside that overlap with the reference. In the experiments, there are significant errors in camera models, so even correctly detected buildings can be displaced from the true positions. To compensate for this, we shift the building positions for maximal overlap with the reference and record the needed displacement as a location error.

8
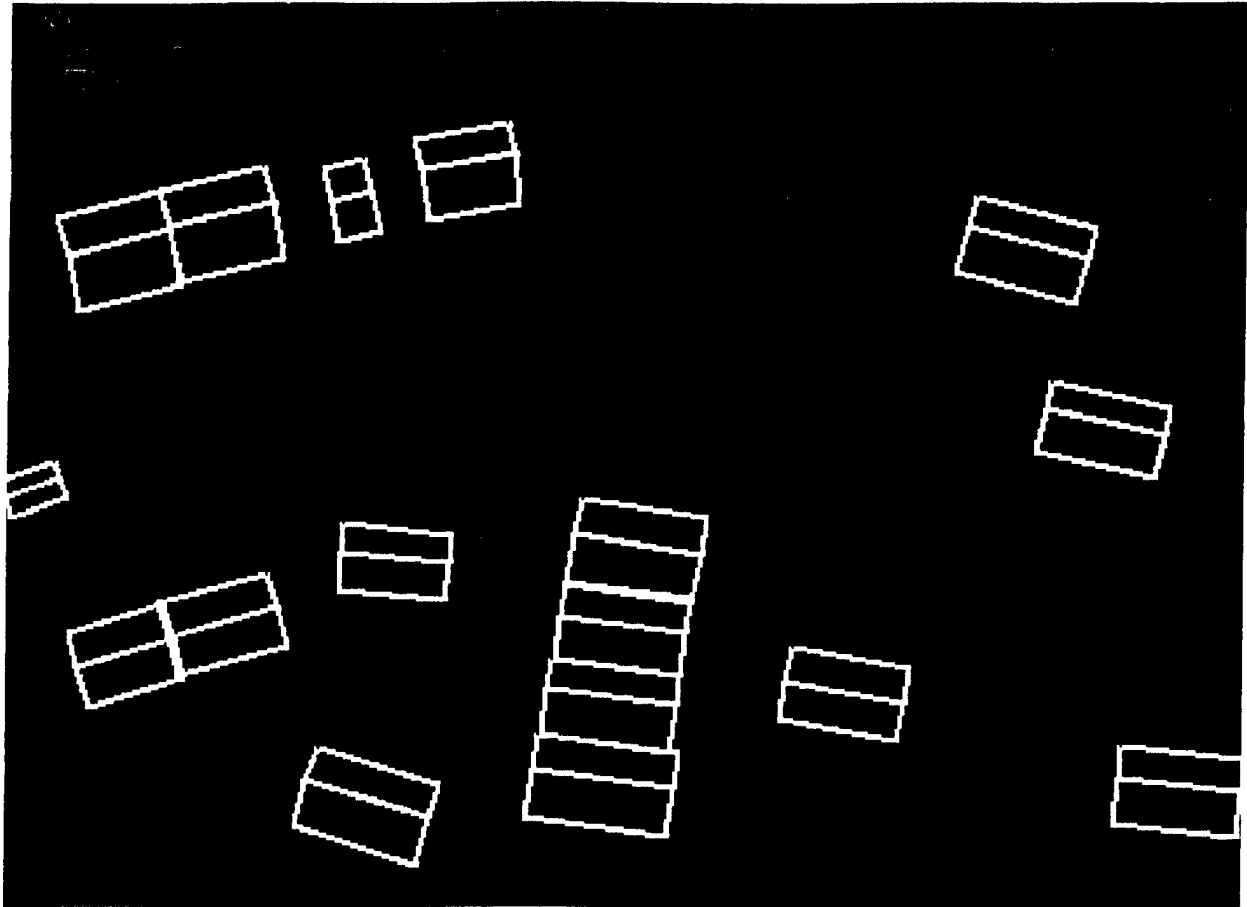
Figure 1.9 Gabled-roof buildings extracted using IFSAR cueing

The procedures to calculate and report performance evaluation figures of our systems are currently under development. We describe some preliminary results for the McKenna MOUT site, in terms of building objects and in terms of areas and volumes, with respect the model shown in Figure 1.11, for both the reference and the detected models. The gabled roofs are replaced by equivalent flat roofs formed by the four corners, so they can be evaluated by our current procedures.

Table 1 shows a summary of detection results for the McKenna MOUT site in terms of objects. Note that, as expected, false alarms disappear when IFSAR cues are available but the detection rate also is slightly lower.

Table 1: Components Evaluation

	PAN Only	With IFSAR
Reference Model	27	
Detected Components	29	25
True Positives (TP)	26	25
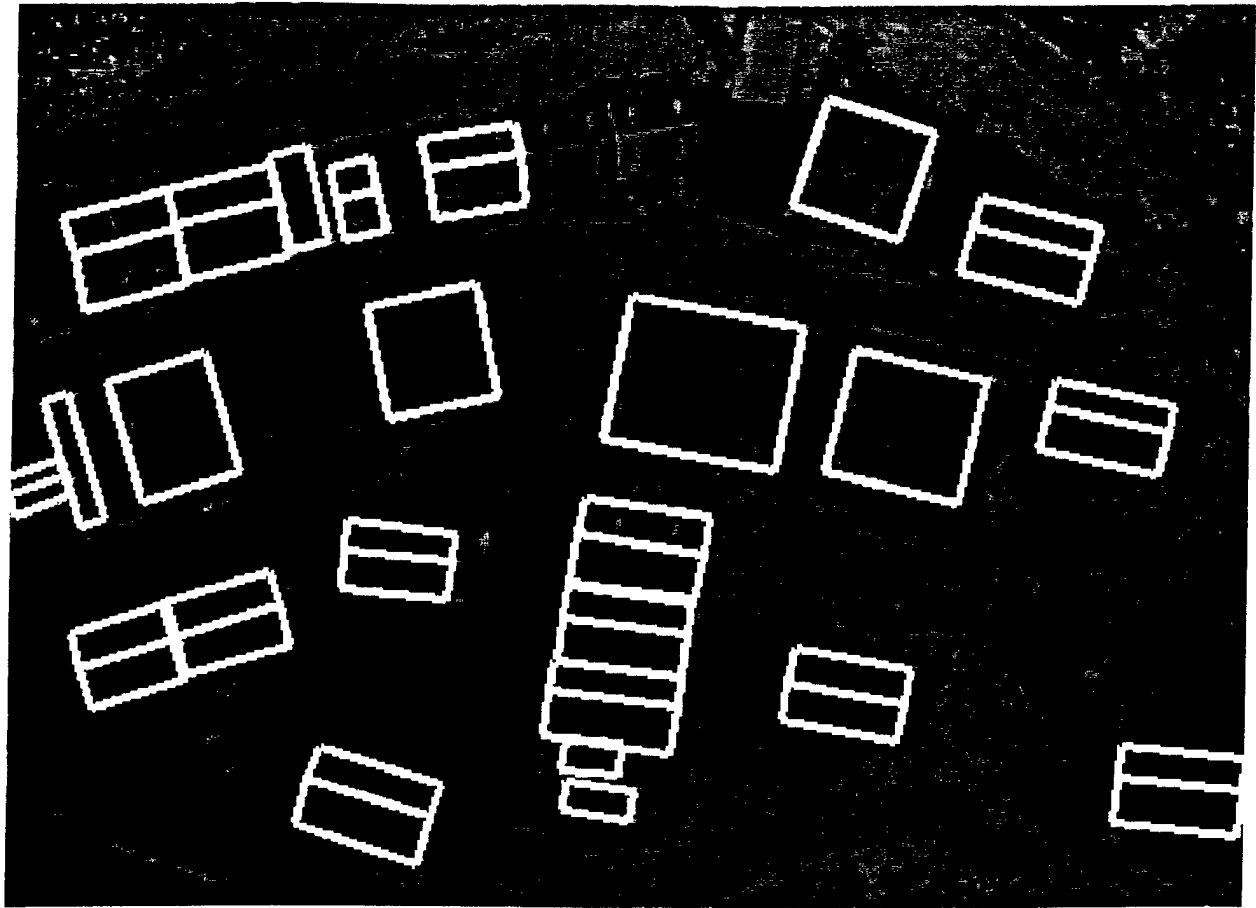
Figure 1.10 Final automatic result for MOUT site using IFSAR cues

Table 1: Components Evaluation

	PAN Only	With IFSAR
False Positives (FP)	2	0
False Negatives (FN)	1	2
Detection Rate	0.96	0.92
False Alarm Rate	0.07	0.00

An important issue is how to combine the results of the above area (or volume) overlap analysis. One can simply consider each area element as an object and count the detection and false alarm rates for all the area elements in the models. Table 2 shows these results for our example (there is no global location error in this case). Ground detection rate is computed for the ground area elements (all elements that are not part of other objects); ground false alarm rate is not shown. Such measures provide some indication of the accuracy of the models but can be misleading. Consider a scene containing one very large and several much smaller buildings. A high detection rate for area elements can be obtained by just detecting the large building accurately and completely missing or poorly detecting all of the smaller ones. This judgement is probably not
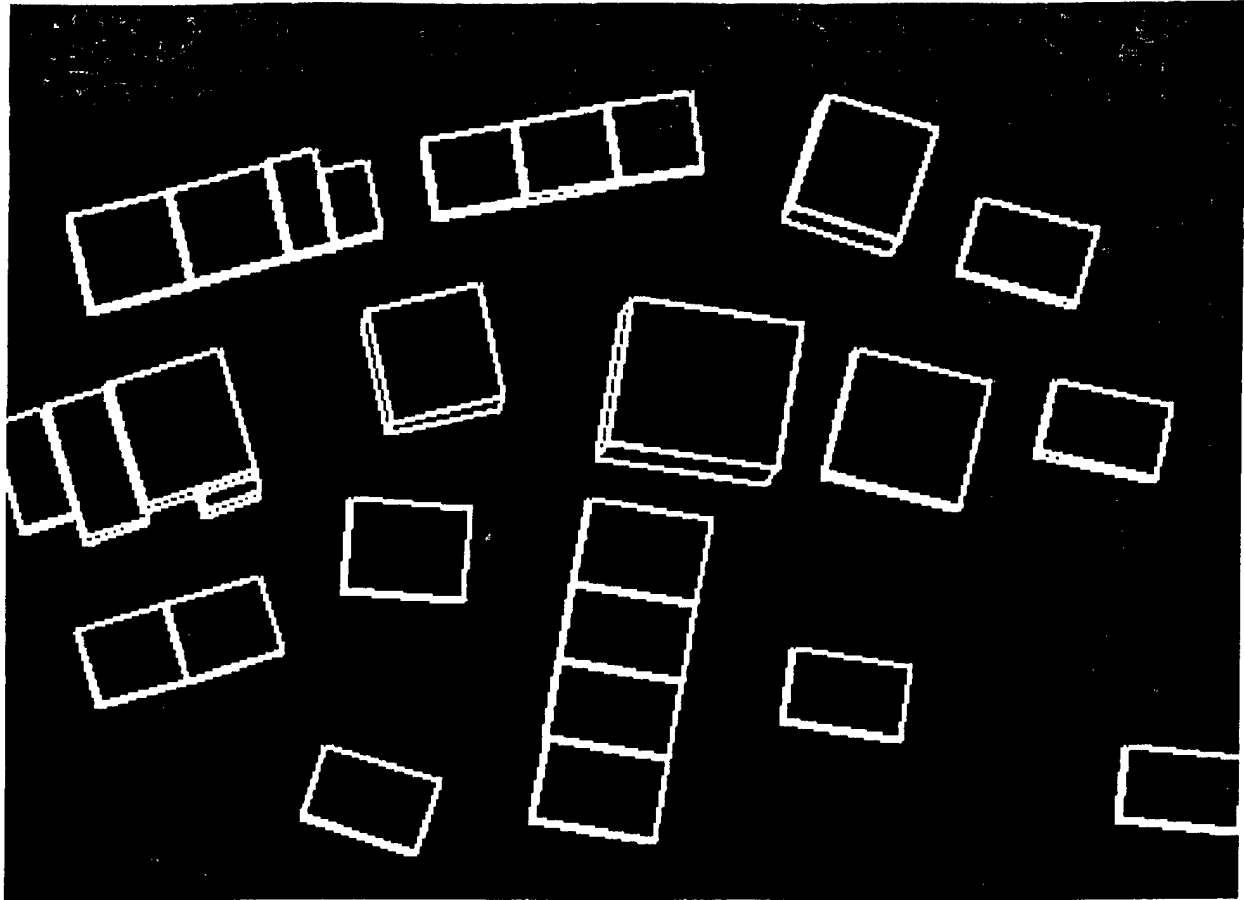
Figure 1.11 Reference buildings for evaluation

consistent with our intuitive judgement and may not be indicative of the utility of the results. Note that a high detection rate for the ground elements can be obtained just by declaring each element to be a ground element.

Table 2: Combined Area Evaluation

	EO Only	with IFSAR
Detection rate	0.8219	0.8341
False Alarm rate	0.1196	0.0407
Ground Detection rate	0.9814	0.9937

To better characterize the accuracy, we compute the detection rates for the area elements of each reference building component and the false alarm rates for each extracted building component separately. When IFSAR cues are available, the performance can be improved significantly. A number of other analyses can be performed on these data. One can, for example, describe how the detection rates are affected by other factors such as building size and volume, scene density, available resolution, etc.

The data in the detection rate tables can be visualized by computing a cumulative distribution of the detection and false alarm rates. Specifically, we can compute the percentage of building components of the reference model whose area (volume) elements detection rate (TP) is at a give value or *higher*. A curve plotting

such a distribution will be called a CDR curve; Figure 1.12a shows the CDR curve for area elements of our example. Similarly, we can compute the percentage of the building components of the extracted model whose false alarm rate (FP) is at a given value or *lower*. A curve plotting such a distribution will be called a CFR curve; Figure 1.12b shows the CFR curve for the area elements of our example.

Figures 1.13a and 1.13b show CDR curves for the volume elements for the reference and extracted building components, respectively. For the purposes of this evaluation, sloping roofs in both the computed and reference models are replaced by flat roofs (to simplify the overlap computation).

A CDR curve that is consistently higher than another CDR curve indicates consistently better performance (similarly, a CFR curve that is consistently lower is consistently better); however, when two CDR or CFR curves intersect, a utility measure, as described in Section 3, needs to be applied to judge which result is more desirable.

(a) CDR Curves
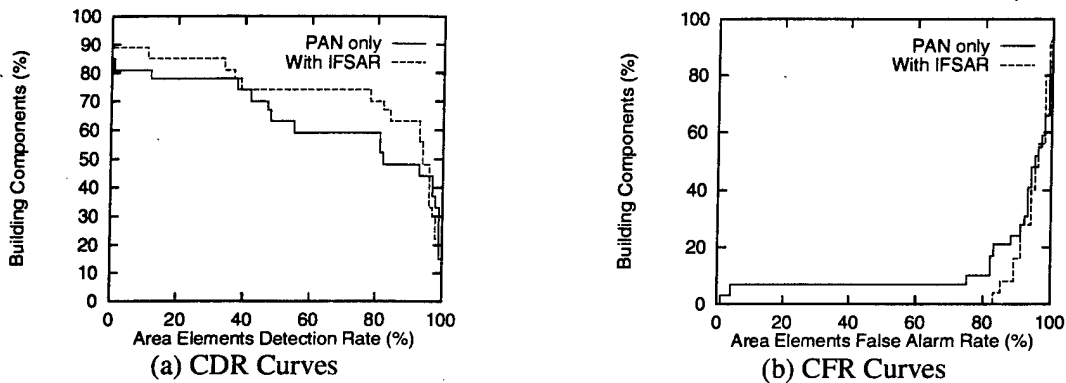
(b) CFR Curves

Figure 1.12 CDR and DFR curves for area elements

(a) CDR Curves
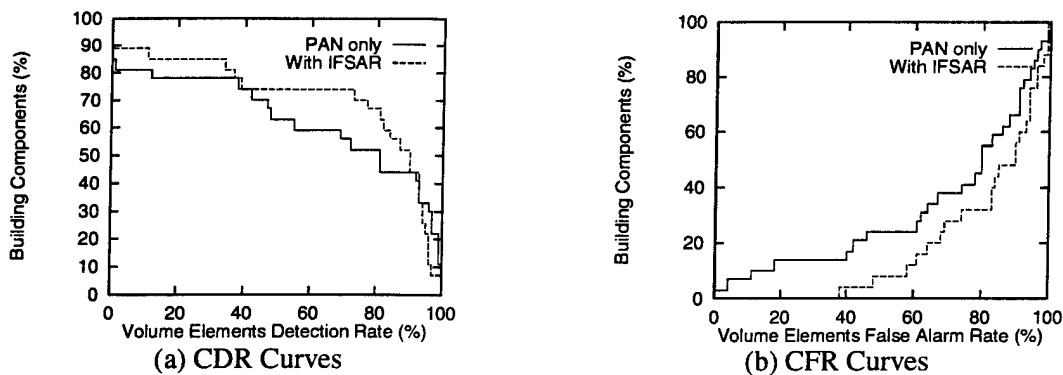
(b) CFR Curves

Figure 1.13 CDR and CFR curves for volume elements

1.4 Interactive Modeling

The results of the automatic system may be edited by a user or the user may provide some cues to the automatic system to improve its performance. The user interaction may occur either after a complete run of the automatic system or only after the stages where lines have been matched and junctions detected. User interactions aid the automatic system in forming new hypotheses. Three-dimensional height computations are still performed automatically. The system requires a user to interact in one image only, even though a second view is displayed and used by the automatic system; this greatly reduces the effort required of the user.

A new building can be added by the user pointing to some corners of the building roof in one image; the pointing need not be precise as the precise corners are selected automatically from the image data. For flat

roofs, between one and three clicks are required. For gabled roofs, between two and four clicks are required. The user given information is used to construct 2-D hypotheses from which 3-D structures are computed automatically. If the computed height is not correct, another single click is usually sufficient to correct it.

Building hypotheses, either derived automatically or by interactions in earlier stages, also can be edited by indicating a corner or side to be changed. More details of this system may be found in [Li et al., 1998].

Figure 1.14 shows the model for the entire Ft. Benning McKenna MOUT site constructed by this procedure (except for one building only partially visible in the window.) The distribution of the needed interactions is given in Table 3. The time measurements apply to user time *after* line and corner matches have been computed and do *not* include the initial set up times. As can be seen, it is possible to construct highly accurate models for a fairly complex site in a very short period of time. In this particular experiment the elapsed time was 165 seconds and only one building required height correction.

Table 3: Distribution of Interactions

Roof type	Clicks needed	Components Formed
Flat-Roof Buildings	1	3
	2	0
	3	4
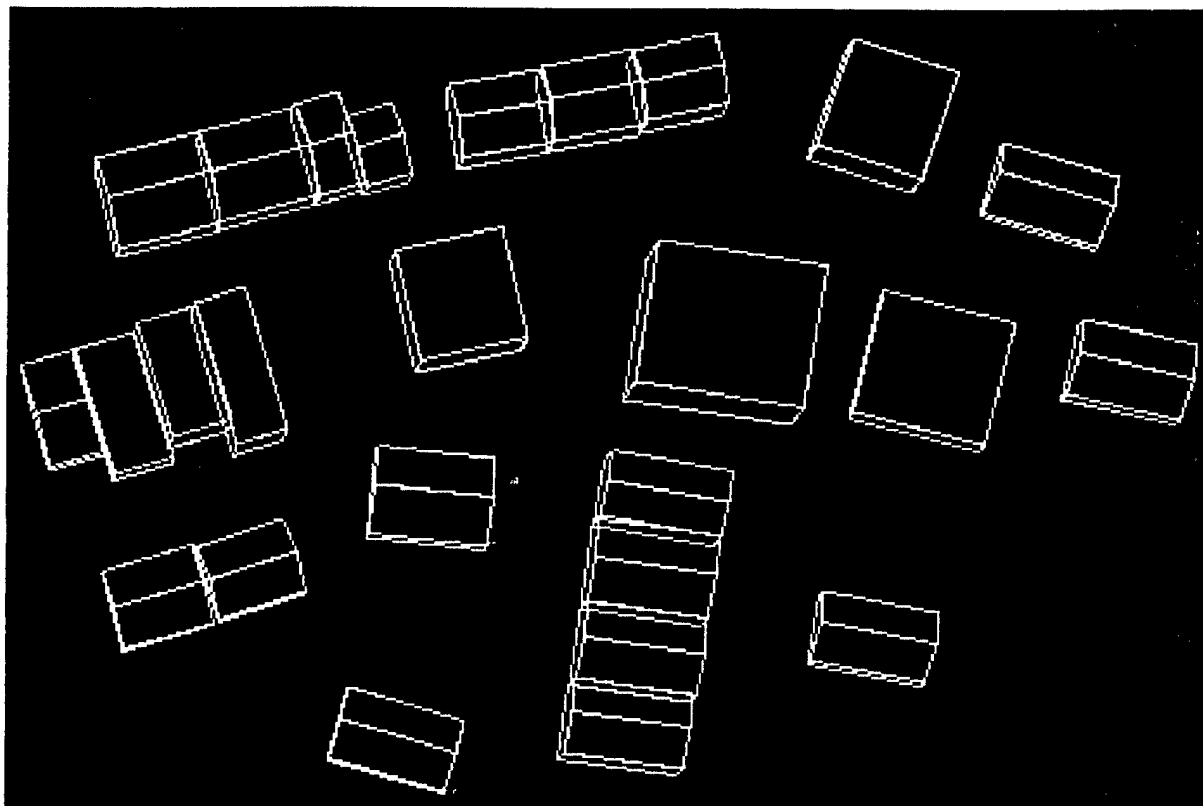
Figure 1.14 User-assisted model of 26 structures takes 165 seconds to construct

13

Table 3: Distribution of Interactions

Roof type	Clicks needed	Components Formed
Gable-Roof Buildings	2	9
	3	8
	4	8
TOTAL	15	26

The user-assisted system is not limited to the underlying capabilities of the automatic system. The collection of processed features (matched lines, parallels and junctions) is available to be used to help model more complex structures. These include buildings with multiple wings, buildings having multiple layers, and structures whose roofs consist of non-rectangular shapes, *i.e.*, polygonal shapes. The system incorporates methods to help construct models of complex structures that include wings and have non-rectangular shapes. The user initiates interaction by adding incremental blocks to a "seed" structure. The seed structure is either generated automatically as a partial detection, or generated manually by the user. The end result of such a process is illustrated in Figure 1.15 (details are given in Section 3). The structures labeled "A" were detected automatically and the buildings labeled "B" were constructed in user-assisted mode by issuing 71 clicks in approximately 55 seconds.
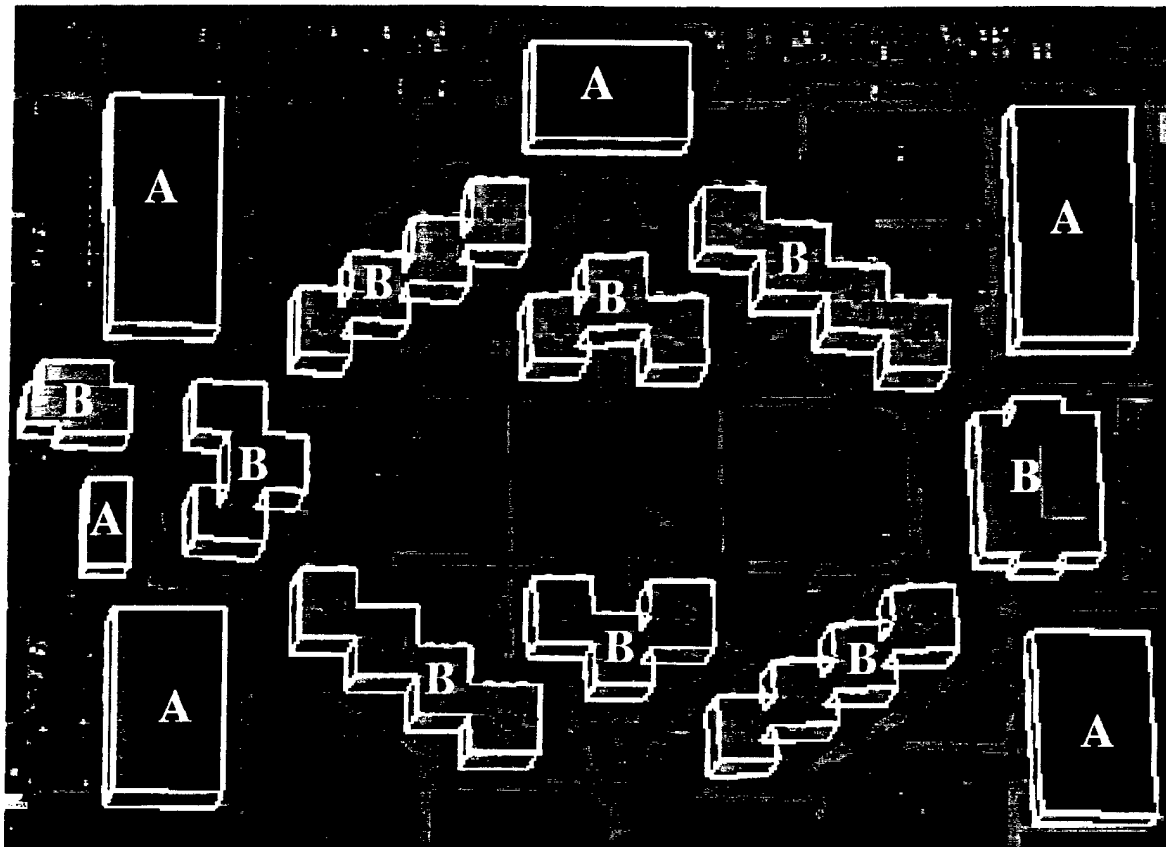
Figure 1.15 User-assisted modeling of complex buildings in Ft. Hood

14

To help evaluate this result consider the number of required clicks compared to the total number of corner points on the modeled roofs; 150 corner points in this example. Thus, at least a 52 percent reduction in effort is achieved. If the number of ground points of buildings (for manual height adjustment) is included, even more reduction is achieved. A comparison between the enhanced capability, the simpler component by component method [Li, *et.al.*, 1998], and a manual method [Strat, *et. al.*, 1992], is given in Table 4.

Table 4: Comparison with other system

Method	Clicks	Time (sec.)
Enhanced Method	71	55.5
Simpler Method	93	77.5
Manual Method	166	676.0

The number of clicks reported, and the time for the other systems shown in Table 4 do not include the number of clicks or the time needed to select items from pull down or other menus, which are extensively used in traditional systems.

2. Use of IFSAR with PAN Images for Building Modeling

In the past, much of the research on automatic building detection and description has been focus on the use of electro-optical panchromatic (PAN) achromatic images [Noronha & Nevatia, 1997, Collins et al., 1998, Roux & McKeown]. PAN images have many advantages. It is relatively easy to acquire resolution images (say on the order of 0.5 meters/pixel). Humans find it easy to visualize such images and to extract the needed information from them; however, their use for automatic extraction has proven to be quite difficult. One of the principal causes of this difficulty is the lack of direct 3-D information in the 2-D images. Three-dimensional information can be inferred for features that can be correctly corresponded in multiple images (assuming knowledge of relative camera geometry) but the correspondence problem is a difficult one as the feature appearances can change in different views and other similar features may exist. Also, aerial images may contain large areas that are homogeneous, such as roofs of buildings where few features exist to match in different views and 3-D information must be inferred by interpolation that requires correct surface segmentation.

In recent years, sensors have been developed that can measure 3-D *range* to a point directly. Availability of this information makes the task of building detection much easier as these structures are elevated above the surrounding background. Two classes of such sensors have been developed. First, LIght Detection And Ranging (LIDAR), uses a laser beam for illumination [Stetina et al., 1994]; distance to a point is determined by the time taken for light to travel to and return from the point (the actual measurement may be done by measuring phase change). The second, IFSAR, computes 3-D position by interferometry from two Synthetic Aperture Radar (SAR) images [Curlander & McDonough, 1991; Jakowatz et al., 1996]. Both sensors use active, focused illumination and rely on reflected light to reach back to the sensor; however, many surfaces act like mirrors at the wavelengths of the respective sensors and those points are not well imaged. Data from range sensors typically have many holes or are even completely erroneous. The resolution of such images is typically lower than that of intensity images. Such images also can be difficult for humans to visualize and fuse with the PAN images.

The complementary qualities of PAN images and range data provide an opportunity for exploiting them in different ways to make the task of building detection and reconstruction easier. We do not propose to combine the two sources at the pixel level but rather to use extract information from each, which is then combined and perhaps used to guide extraction of additional information. In particular, we feel that the range data are suited for detecting possible building locations and in making coarse models for them while the PAN data are suited for confirmation and for accurate delineation and description. Other approaches of range data to use may be found in [Hoepfner et al., 1997; Chellappa et al. 1997; Haala & Brenner, 1997].

Our baseline system for extracting buildings from PAN images is described earlier in Section 1 [see also Nevatia & Huertas, 1998] and in [Noronha & Nevatia, 1997]. The approach is basically one of hypothesize and verify. *Hypotheses* for potential roofs are made from fragmented lower-level image features. The system is limited to buildings with rectilinear shapes causing the rooftops to be parallelograms, or combinations thereof, in an image. Promising hypotheses are *selected* among these by using relatively inexpensive evidence from the rooftops only. The selected hypotheses are then *verified* by using more reliable global evidence. The verified hypotheses are then examined for overlap which may result in either elimination or in merging of them. The methodology is to be liberal at the early stages, and make decisions only when sufficient information is available to make them reliably. The system described in this report differs from the previous ones in its use of a Bayesian classifier for decision making at the various stages.

The performance of this system can be improved greatly by using range data in the process. Range information can be useful at all stages of the system: in hypothesis formation, selection, and verification. At the early stages, the range data can limit the region of interest and provide cues for the kinds of hypotheses to construct. At the later stages, range data can be used, along with heights computed by stereo correspon-

dence, to increase (or decrease) the confidence in whether a hypothesis actually corresponds to a building, and the reliance on shadow evidence can be reduced.

In the next section, we describe how useful *cues* can be extracted from the range data. Use of these cues in the building extraction process is then described. Results comparing the effects of these cues are presented in Section 2.2.

2.1 Cues from Range Data

We aim to extract cues for presence of buildings. When the range data are of high quality such as for digital elevation models (DEMs) derived from stereo PAN images (possibly including some manual editing) or by a high resolution IFSAR searchlight IFSAR sensor, the process of extracting cues is rather simple. It consists of extracting the appropriate regions from the convolution with a Laplacian-Of-Gaussian filter (LOG), and selecting those having appropriate size.

Next, we discuss the use of low resolution IFSAR data that are much more common (such as from the IFSARE sensor). IFSAR data are usually given in the form of three images, called the **mag, dte,** and **cor** images. The **mag** image is like a normal intensity image measuring the amount of reflected signal coming back to the sensor. The **dte** image encodes the 3-D information in the form of a digital terrain elevation map where the pixel values define the height of the corresponding scene point. The **cor** image contains the phase correlation information between two images used for the interferometric process; it can be useful in distinguishing among types of materials as the returns associated with objects that remain stationary are highly correlated.

Although the primary source of cues for buildings appears to be the **dte** image, an initial characterization of the data indicates that at low resolutions it is appropriate to extract cues that indicate the possible presence of significant features, such as buildings and trees, from a combination of the **mag, dte,** and **cor** images. Consider the portion of an image from the Fort Hood site shown in Figure 2.1 It contains 11 buildings. The corresponding 2.5 meter **mag, dte,** and **cor** images are shown in Figure 2.2. Only some buildings appear salient in the **dte** image and most buildings are represented in the **mag** image.
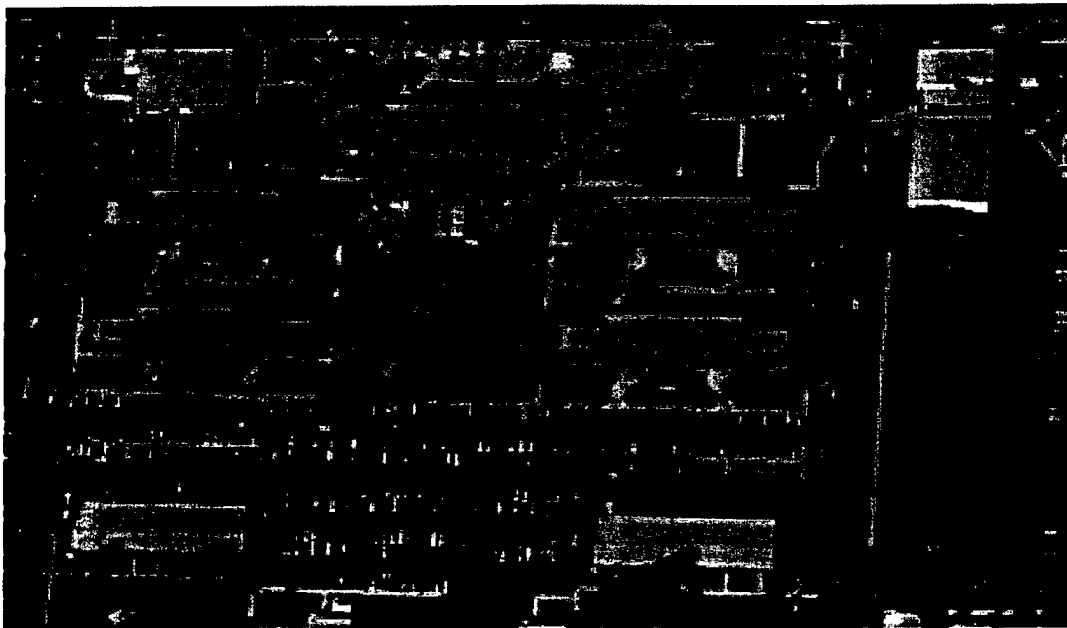

Figure 2.1 Portion of Ft. Hood PAN image

17

(a) mag image

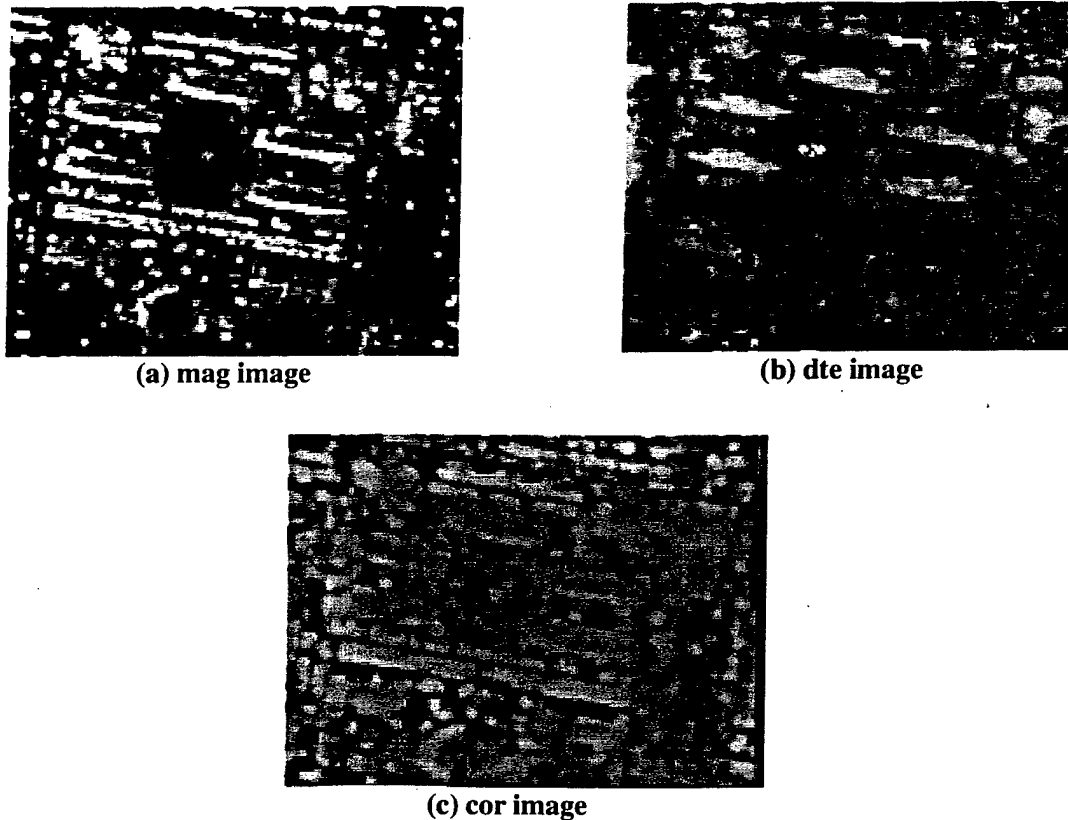(b) dte image

(c) cor image

Figure 2.2 Image components from IFSARE sensor

Clearly the orientation of the buildings with respect to the direction of flight during SAR image acquisition contributes to the resulting returns, and some buildings may not be adequately represented for some orientations. It is advantageous to use a combination of the **dte** and **mag** images with a higher weight given to the **dte** contribution. The phase correlation image is used to take advantage of the behavior of the signal phase in the IFSAR process, thus helping support cues for buildings derived from **mag** and/or **dte** components taken individually or in combination.

Figure 2.3 shows that, at these low resolutions, it is not sufficient to threshold the **dte** images to obtain cues corresponding to objects of interest. Figure 2.3 (left) shows the **dte** regions "just above the ground," that is, about 1.5 m above the ground (mean elevation). Figure 2.3 (right) shows the thresholded **dte** image at the mean intensity plus one standard deviation. The buildings are somewhat apparent in this image but the presence of many artifacts would be misleading to an automated system beyond a rough indication of possible presence of a building.

Instead, the **mag** and **dte** images are processed by applying a technique similar to the one described in [Chen, et al., 1987]. Regions of interest are extracted from the **mag** and **dte** images, and correspond to the positive-valued regions in the output of the convolution of the images with a Laplacian-of-Gaussian (LOG) filter. These images are combined linearly with a weight of 10 given to the **dte** regions. The resulting image is then thresholded to determine the regions of interest. The *cues* image, shown in Figure 2.4 (left) is computed by selecting those regions that have a high correlation component. Figure 2.4 (right) shows the connected components that have a certain minimum size (area) and are taken to correspond to cues for building structures and other tall objects. Note that all the buildings are well represented except for the one in the lower left, and the one on the lower right, which is not represented.

18

Figure 2.3 Thresholding of dte image. At mean (left); at mean plus one standard deviation

Figure 2.4 Computed cue regions (left) and cues selected by size (right)

2.2 Integration of Cues into the Building Detection System

We next describe the use of these cues in the multiview building detection and description system described in [Noronha & Nevatia, 1997]. This system has three major phases: hypothesis formation, hypothesis selection, and hypothesis validation. Range cues can assist the process at any or all of the three stages as discussed below.

Hypothesis Formation:

Cues can be used to significantly reduce the number of hypotheses that are formed by only considering line segments that are within or near the cue regions. As many false hypotheses are eliminated, the hypotheses formation can be made more liberal to include some hypotheses that may have been missed otherwise. The set of linear features, shown in Figure 2.5 for one of the three views processed, represents a reduction of 95.6 percent when IFSAR cues are used (the number of hypotheses formed is reduced by 48 percent).

Hypothesis Selection and Verification:

For each hypothesis, support from IFSAR analysis is calculated. The hypotheses is projected onto the IF-SAR image and overlap of the projected roof with IFSAR regions is computed. The current system requires that the overlap be at least 50 percent of the projected roof area.

Figure 2.6 and Figure 2.7 show selected and final hypotheses, with and without the use of the IFSAR cues, respectively. Note that the building on the lower right is not found (Figure 2.7, bottom) as the lack of a cue

19
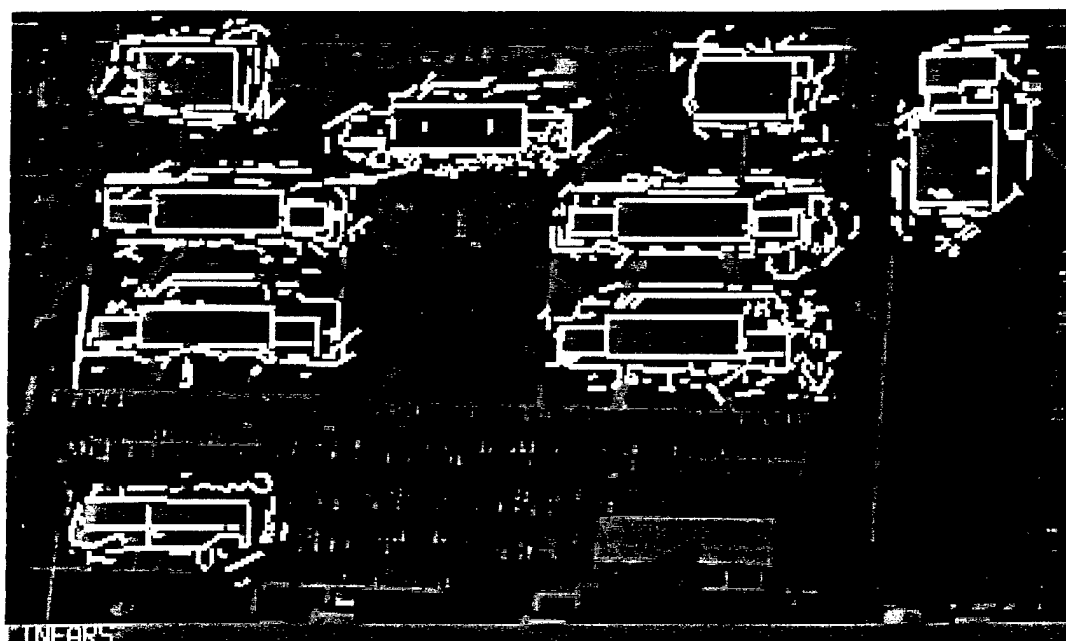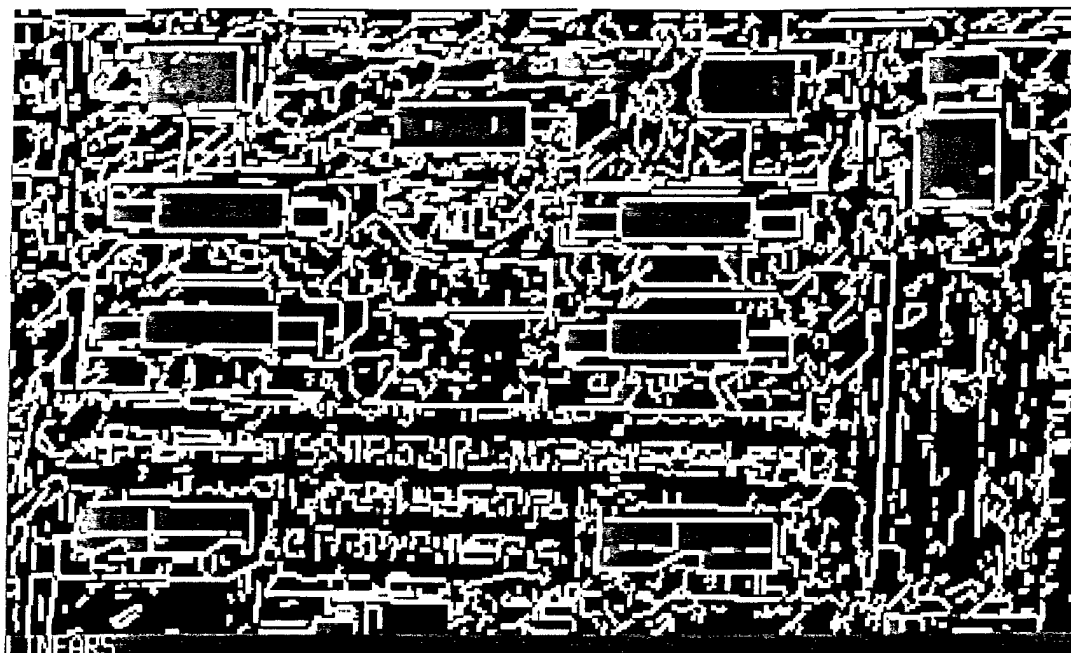
Figure 2.5 Linear segments in PAN image (top) and those near IFSAR cues

prevented a hypothesis to be formed there. On the other hand, the building on the middle top is not found without IFSAR support (Figure 2.7, top) but found with it (Figure 2.7, bottom).

Table 7 gives a comparison of the number of features and final result counts with and without use of IFSAR cues. Section 2.3 provides evaluation details for the quality of the results obtained.

20

Figure 2.6 Selected Hypotheses using PAN only (top) and using IFSAR cues

21
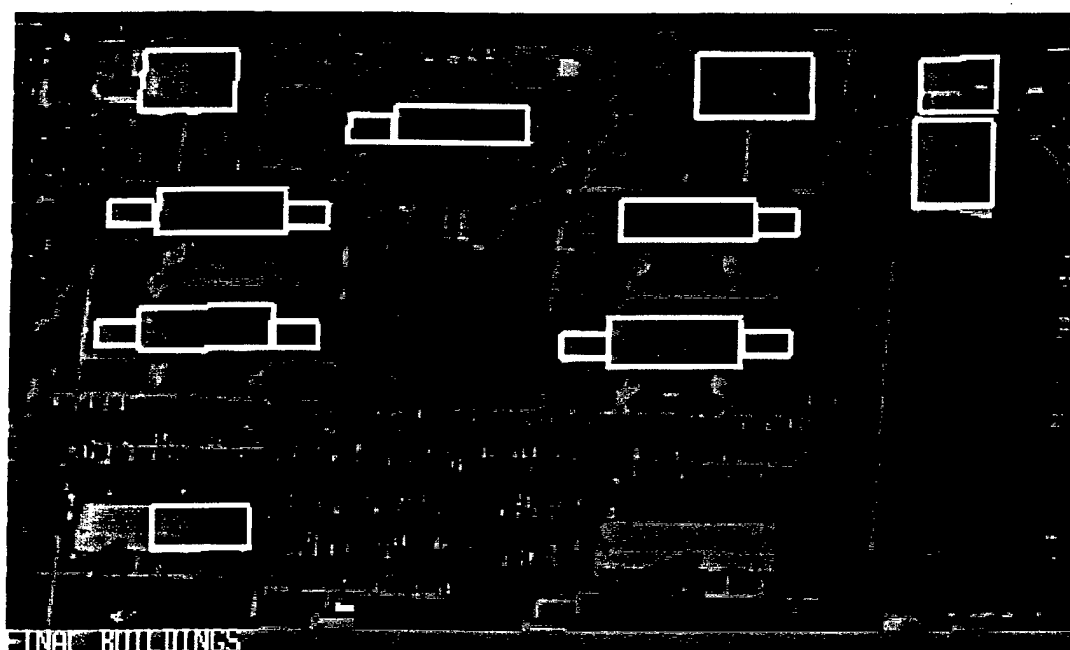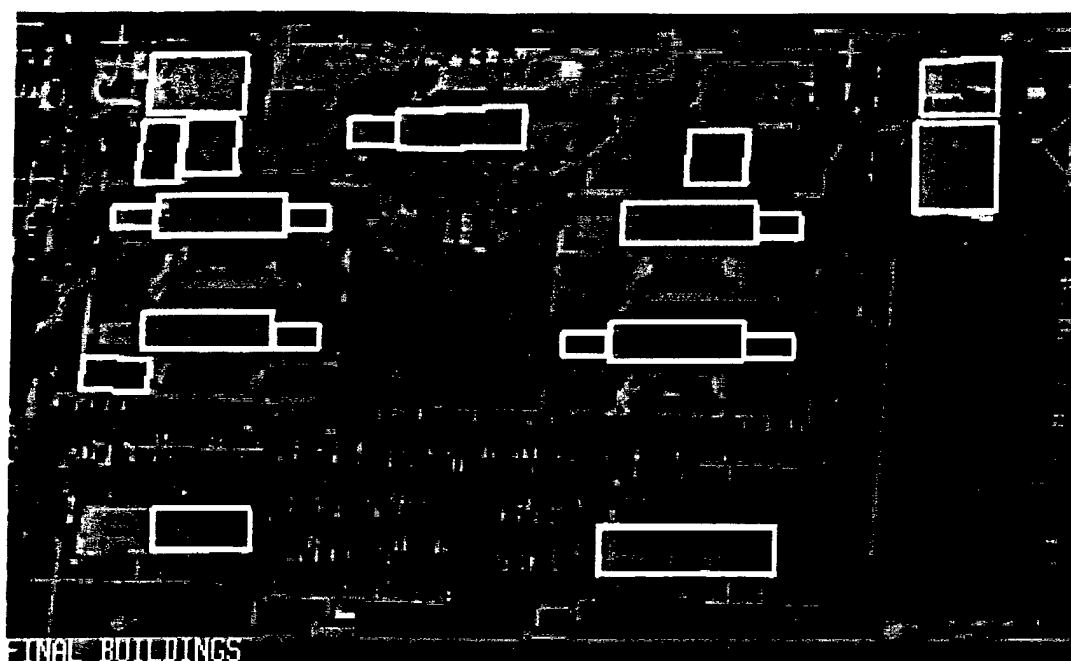
Figure 2.7 Final hypotheses using PAN only (top) and using IFSAR cues

22

Table 5: Automatic Processing Result

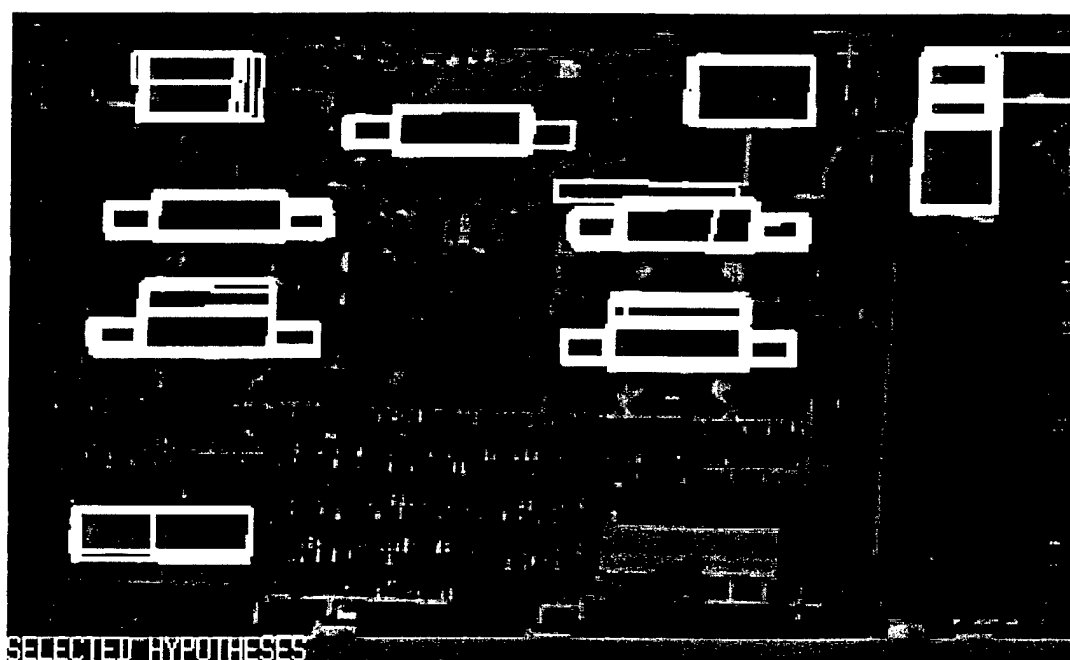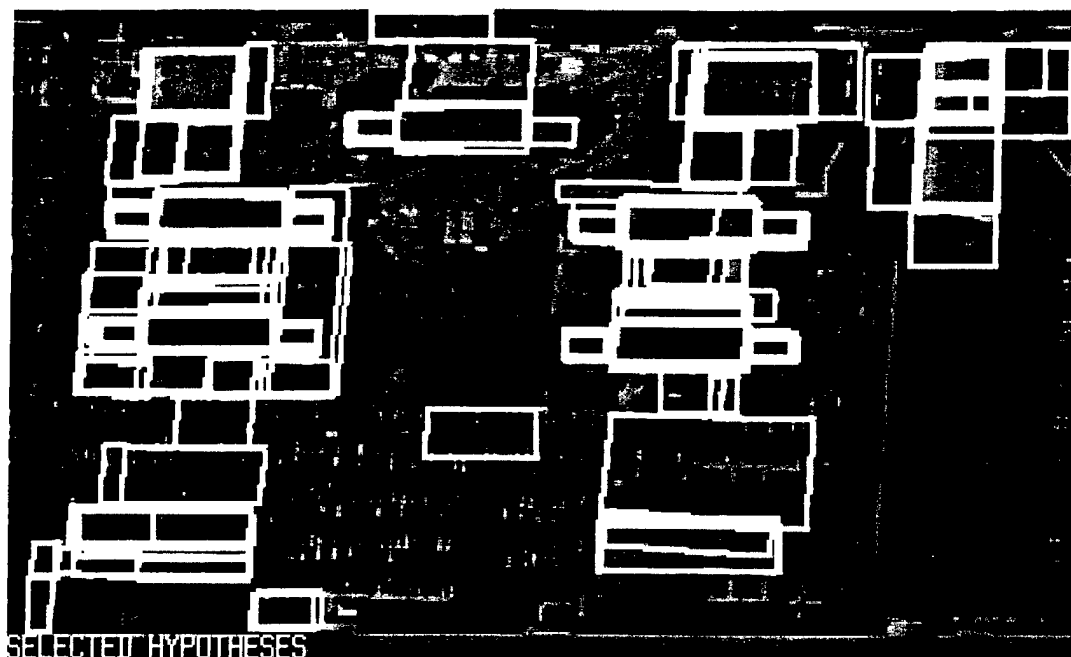| Feature	PAN Only	With IFSAR
Line segments	7799/7754/5083 (from three views)	
Linear structures	2959/2963/2042	669/650/552
Flat hypotheses	2957	1732
Selected hypotheses	383	296
Verified hypotheses	215	192
Final hypotheses	21 (4 false)	18 (0 false)

2.3 System Evaluation

We use the methodology described in Section 1 to evaluate the results of the system with and without use of IFSAR. The evaluation uses detection rate and false alarm rate measures defined in Section 1.3. We consider a building to have been detected if *any* part of it has been detected. The amount by which a building has been correctly detected is computed by the number of points inside that overlap with the reference.
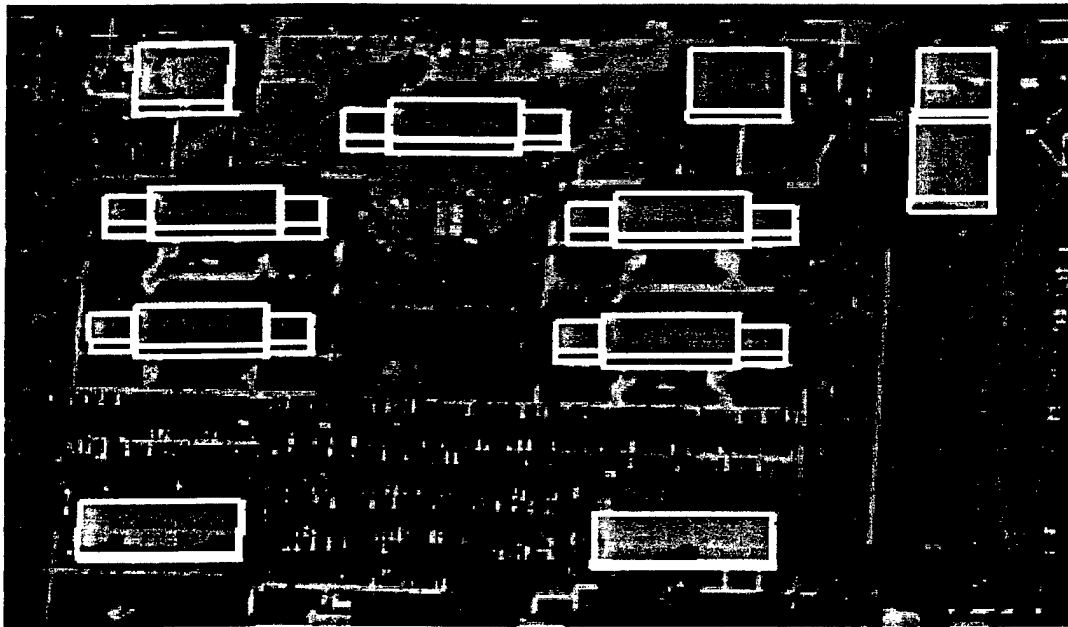
Figure 2.8 Reference model for evaluation

Table 8 shows a summary of detection results for our Ft. Hood example in terms of object components. Note that false alarms disappear when IFSAR cues are available.

Table 6: Components Evaluation

	PAN only	With IFSAR
Model	21	
Total	21	18
TP	18	18
FP	4	0
FN	2	1
Detection Rate	0.90	0.95
False Alarm Rate	0.18	0.00

The combined results in the form of CDR and CFR curves (see Section 1.3) are shown in Figure 2.9, for the *area* analysis, with and without IFSAR cueing, and Figure 2.10 shows a similar graph for the *volumetric* analyses.
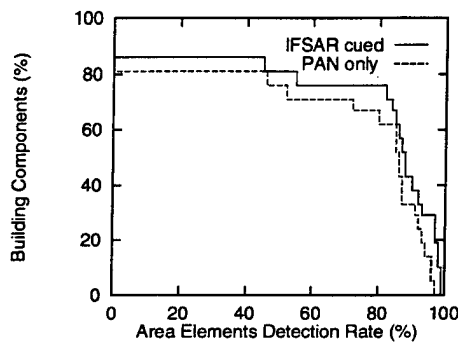
(a) CDR Curve
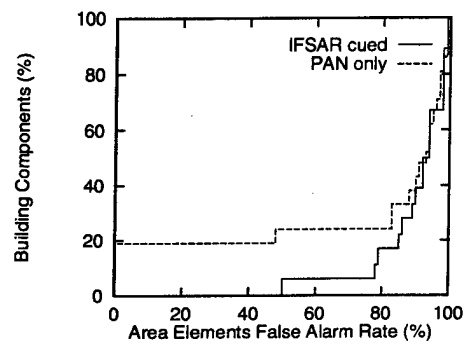
(b) CFR Curve

Figure 2.9 Evaluation curves for area elements
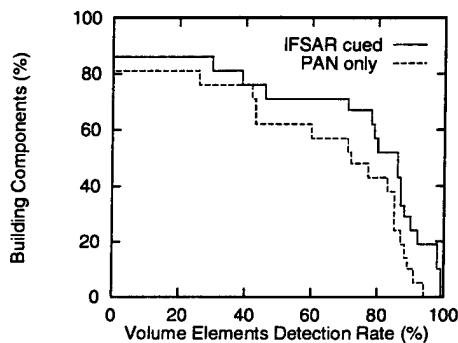
(a) CDR Curve
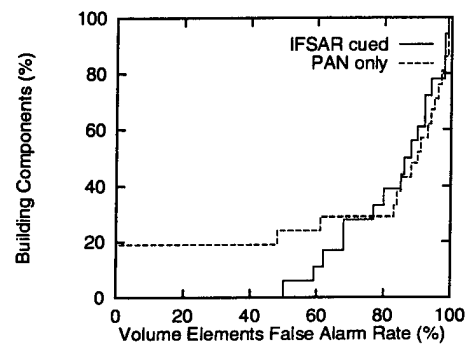
(b) CFR Curve

Figure 2.10 Evaluation curves for volumetric elements

24

Table 7 shows the combined area analysis for the results shown in Figure 2.7. The ground detection rate is the percent correct classification of ground pixels. When IFSAR cues are available, the performance can be improved significantly. A number of other analyses can be performed on these data. One can, for example, describe how the detection rates are affected by other factors such as building size and volume, scene density, available resolution, etc.

Table 7: Combined Area Evaluation

	PAN Only	with IFSAR
Detection rate	0.7461	0.7545
False Alarm rate	0.2023	0.0883
Ground Detection rate	0.9688	0.9863

3. User-Assisted Modeling of Buildings

Generating 3-D models from a set of images is a common task for computer vision. In spite of substantial research in this area, the performance of machine algorithms remains significantly below that of humans. In this work, we explore an approach to bridge this gap by allowing a human "in the loop" but by requiring only simple interactions from the user to generate accurate models efficiently. This approach is illustrated for the task of building modeling from aerial images, which is a difficult and important task.

Significant progress has been made in recent years in the goal of extracting models of buildings from aerial images by completely automatic systems [Grün and Nevatia, 1998; Grün and Baltsavias, 1997] but the results are not completely accurate. Completely manual systems require an unacceptable amount of effort from a human modeler, both in terms of time and cost. We describe an approach that attempts to provide user assists to an automatic system in a way that the user effort is diminished significantly while the quality of the results is still preserved.

Several approaches to user assisted modeling are possible. The conventional approach is to provide a set of generic models that are then fit to the image data by changing model and viewing parameters [Strat et al., 1992]. In this approach, the system provides geometric computations but substantial time and effort are required from the user. Newer approaches have attempted to combine user input with varying amounts of automatic processing. In [Heuel and Nevatia, 1995], the authors suggest providing only an approximate building location to extract a building. In [Hsieh, 1996], other interactive tools are described including methods for replicating model buildings that are identical or very similar to others. In [Grün and Dan 1997], an automatic system constructs topological relations among 3-D roof points collected by a user for each roof; this system can work with several types of complex roofs.

In our approach, basic modeling tasks are still performed automatically, and this system receives simple, but critical, assists from the user. The assisted system's capabilities are limited by those of the underlying system. In this case, the shapes of the buildings are restricted to be rectilinear; the roofs may be either flat or symmetric gables.

The underlying automatic system is the USC Multiview Building Detection system described in [Noronha & Nevatia, 1997]. The basic steps of this system consist of forming parallelogram hypotheses (to represent rectangular parts of roofs) in one image (but by using information about matched lines) and inferring 3-D shapes from them. Three-dimensional analyses also provides verification of the hypotheses by determining if sufficient evidence for their presence exists. A user can assist this system in the process of hypotheses formation as well as in making corrections to the resulting 3-D models.

A user interaction typically consists of the user pointing to a point or line feature; the pointing need not be precise as precise features are automatically selected by the system. We will call one such interaction a "click." The system requires two (or more) views of a scene with associated camera geometry; however, all user interactions take place in one view only. Other views can be displayed but the user is not asked to view the images stereoscopically. We believe that confining interactions to one view can significantly reduce the effort required by the user.

The following system description is divided into two components. In the first situation a building has not been detected and needs to be added. In the second, a building has been detected partially and requires editing.

3.1 Adding a Building

User assistance in adding a new building to the model consists of approximately indicating some numbers of corners of its roofs. The automatic system constructs and displays a 3-D model after each click (though in some cases, no models may result). Each subsequent click refines the models. For each rectangular com-

ponent of a rectilinear building, up to three clicks may be required for flat-roofed buildings and up to four clicks for gabled roofs. The resulting models can then be edited by the methods described in Section 3.3.

After each click, indicating the approximate position of a corner, the system finds nearby corners constructed from extracted matched line segments. The sides of a single corner can suffice to trigger some parallelogram hypotheses. Multiple hypotheses are possible at each step; selection among them is made by searching for matching evidence in all the available views. The matching process also creates 3-D models from the initial 2-D roof hypotheses.

The operational and computational processes for the flat and the gabled roofs are quite similar, however, for simplicity, we describe each of them separately below.

3.1.1 Flat-Roof Buildings

A maximum of three corner clicks is required to specify a rectangular flat roof component.

3.1 depicts the situation after the first user click. The system locates all junctions near the click and reports a failure if none is found. For each junction found, the system attempts to construct a parallelogram. The parallelogram is formed by first examining the stored information and looking for a U-structure that uses the junction legs. If no U-structure is available, the junction legs are used to derive the parallelogram (roof hypothesis). The elements of the parallelogram are matched to elements on the other views and scores are computed as the system would during automatic operation. The system then selects one configuration and presents it to the user.
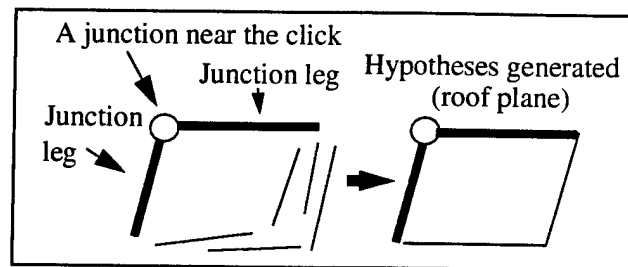

Figure 3.1 First input analysis

3.2 illustrates the situation after the second click. The second click is used to generate a new hypotheses in the same manner as with the first click. The hypotheses are formed that include the point from the first click, however, they are weighted higher.
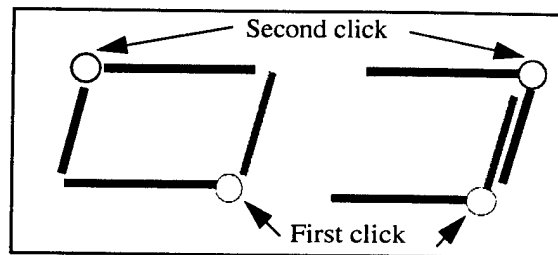
Figure 3.2 Two possible configurations for a second click

After the third click, three points are used to form three possible parallelograms to represent roof hypotheses, as shown in Figure 3.3. The system calculates the 3-D orientation of these planes and matches the elements with elements on other views. For all possible matches, select those hypotheses with least inclination for a flat-roofed building. Also the angles between the sides must be close to 90 degrees in 3-D. The system computes scores as before, and selects the hypothesis with the best score.

27

Figure 3.3 Three parallelograms can be formed from three points

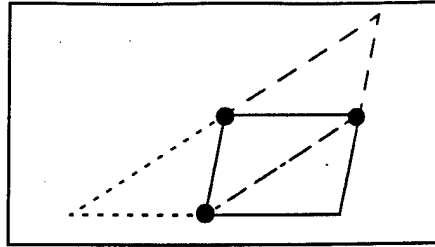To convert a 2-D roof hypothesis to a 3-D building hypothesis, we need to compute the applicable height for the roof hypothesis. This height could be computed by using matches for the linear features forming the hypotheses in two or more views; however, we have found it more robust to simply conduct a search for different height values in a range by small steps. For each hypothesized height, the 3-D model is projected to a second view and supporting evidence for it, consisting primarily of nearby and overlapping linear features, is collected. Only the roof edges are considered. The height providing the best score is selected. The height of the ground is assumed to be known (or can be computed from a separate user interaction).

We show some examples of user interaction on small windows. Figure 3.4 shows two examples where a single click was sufficient to recover each of the two buildings, with no further editing required.
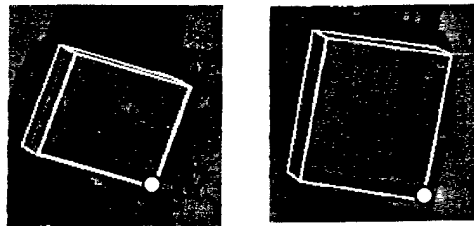

Figure 3.4 Addition of two buildings by one click each

Figure 3.5 shows an example where three clicks are needed. The first click results in a partial hypothesis. The second click gives a better hypothesis but is still not completely accurate. The third click results in an accurate model that requires no further editing. For each of these examples, the time required to construct a model is less than 4 seconds.
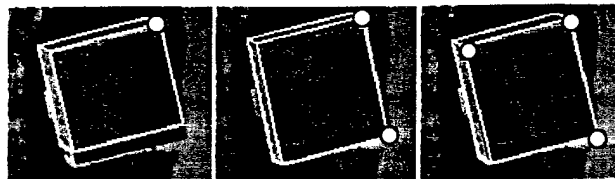
Figure 3.5 Addition of a building by three clicks

3.1.2 Gabled-Roof Buildings

At least two corner clicks but not more than four corner clicks are required to generate a rectangular symmetric gabled roof component.

First, the user clicks on or near two corners on the roof spine of the desired building. The system locates all the junctions near these two clicks. If none is found for either corner, the system reports a failure and prompts the user to give more clicks. Otherwise, for each junction found, the system tries to find all slanted side edges that make right angles with the spine edge. Among all these slanted side edges, the system eval-

28

uates their goodness and picks a best one; thus we can obtain a derived junction for one side of the roof. The situation after the first two clicks is shown in Figure 3.6.
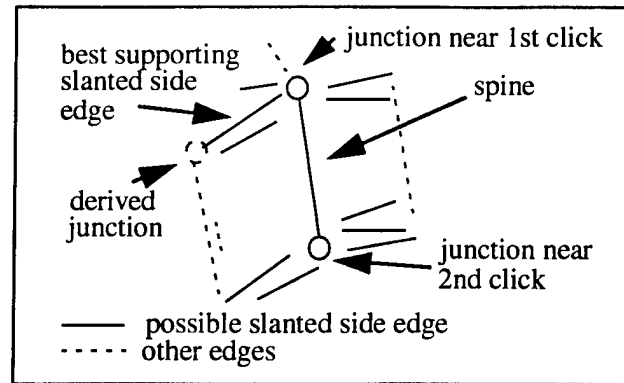

Figure 3.6 The situation after first two clicks

If the above derived junction cannot be obtained or is inaccurate, the user may click on or near a corner of the roof side. With these three junctions obtained either by clicks or from computations, the system then tries to locate the symmetric corner of the currently obtained side corner. Because the side height is not known yet, the system assumes a medium height for this corner and then projects it from the image coordinate to the world coordinate. Using symmetry property, the system is able to acquire the 3-D coordinate of the reflected corner. The system then projects it back to the image coordinate, locates all possible corners around, and chooses a best one as the other derived junction. Figure 3.7 illustrates the use of symmetry property in obtaining the other derived junction.

**Figure 3.7 Use symmetry property to try to compute
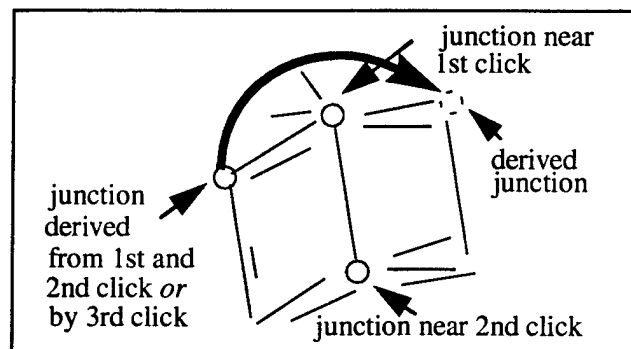the symmetric junction**

If the other derived junction is inaccurate, the user needs to provide a fourth click. With all these four junctions obtained either by clicks or from computations, the shape of the gable roof is determined and the corresponding 2-D roof hypothesis can be generated, as shown in Figure 3.8.
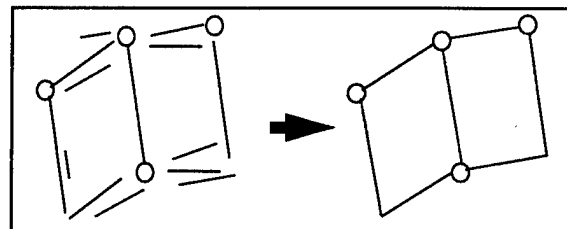

Figure 3.8 2-D Gabled roof hypothesis formed

29

To form a 3-D building hypothesis, we need to compute the heights of the side and the spine of the gable roof. This is done by a process similar to that for flat roofs and consists of searching in the space of allowed side and spine heights, projecting the 3-D hypotheses to another view and collecting the supporting matching evidence. As this method requires a search over two parameters, a two-level matching analysis is used to speed the computations. In the first level, only the spine height is varied. For each spine segment, we compute its score as the accumulated matching spine edge evidence in another view. Only those segments with scores higher than 70 percent of the highest score are kept for continued matching analysis. In the second level, we change the side height by small steps in the allowed range. For each side segment and each remaining spine segment, we project the hypothesis into another view and compute the score as the accumulated matching roof edge evidence. The pair of side height and spine height that scores best in the second level are used to convert 2-D hypotheses into 3-D hypotheses.
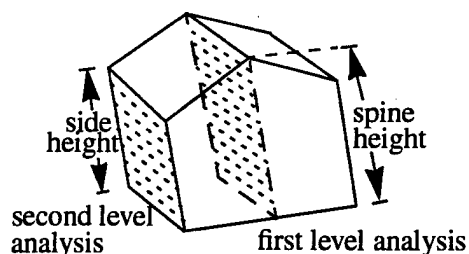

Figure 3.9 Two-level analysis of height for gabled buildings

We show some examples of user interaction for the Gabled-buildings. Figure 3.10 shows two examples where two clicks are sufficient to detect symmetric gabled buildings, with no further editing required.

Figure 3.10 Addition of gabled buildings with two clicks

3.11 shows an example where a third click is needed to refine the hypothesis from the previous two clicks. The third click generates the accurate model that requires no further editing.
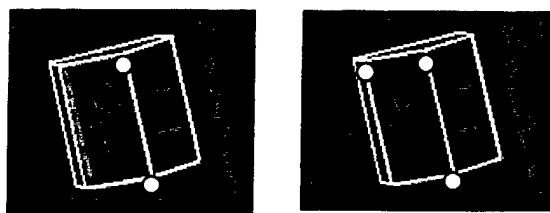
**Figure 3.11 Addition of a gabled building where the 3rd click
is needed to refine the hypothesis**

Figure 3.12 gives an example where four clicks are necessary. The system finds no hypothesis with the first two clicks. The third click results in an inaccurate hypothesis; thus, the last click is essential to form the accurate hypothesis. For each of the above gabled buildings, the time to construct a complete model is less than 6 seconds.
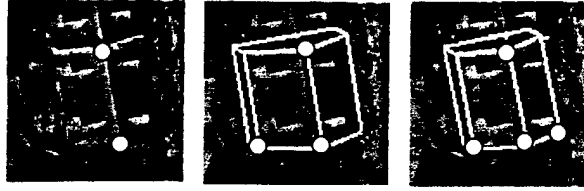
Figure 3.12 Addition of a gabled building where four clicks are needed

3.2 Editing a Building

This process can be used to edit a current building hypothesis, either derived automatically or by interactions in earlier stages as described above. Three available corrections are to adjust a corner, to adjust a side or to adjust height. If more complex interactions are needed, the building should be deleted and reconstructed. Again, interactions take place in only one view and only approximate locations are needed. The user interactions are as follows:

- To adjust a corner, click on a new location for it,
- To adjust a roof side, click anywhere along the actual roof side, and
- To adjust height, click anywhere along the base of the building.

For each of the steps above, the system recomputes all aspects associated with the formation of 3-D hypothesis during automated operation. The actions of the automatic system for editing a building are similar to those for adding a building. If a corner is indicated by the user, the system finds the nearest corner in the existing hypotheses and replaces it by a corner near the indicated position. A new hypotheses is generated and its height recomputed. If a side is indicated, the closest side of the existing hypothesis is found and moved to include the indicated position. Again, a new hypothesis and building model is constructed.

Correction of height is more complex. A wrong height is obtained by wrongly matching lines during height computation. In the view that the user interacts with, a building with the wrong height will still appear as having its roof in the correct place but the baseline will be wrong (in the other view, either the base or the roof or both may be wrong). As illustrated in Figure 3.13, the roof corner is correct in 2-D, but incorrect in 3-D because of its wrong z-coordinate inferred from automatic height calculation. The user may correct his error by clicking anywhere along the correct base of the building. The system first calculates the correct ground corner by intersecting the new base line (parallel to a roof line) with one of the projected walls. If we use this ground corner and the current hypothesis, the roof corner will project along one of the wall lines but not at the correct roof corner position, also shown in Figure 3.13. We now need to correct the height so that the projected corners and the roof corners coincide. Note that this can not be done simply by scaling the projected line lengths as the projection does not necessarily preserve length rations. For simplicity, we find the correct height by a binary search. During each iteration of the binary search, the system takes the middle height in current height range and uses it to calculate the 3-D coordinate of the roof corner and projects it back to 2-D. The error between the projected 2-D value and the correct 2-D value tells the system the direction in which the height range should be reduced. This iteration is repeated until certain accuracy is reached or the number of iteration exceeds the maximum number of iterations allowed. In the latter case, the system indicates that the clicked ground point is not acceptable.

These processes are illustrated by examples below. 3.14 shows a detected building that is only partially correct. The appropriate correction consists of indicating a point along the actual building boundary to cause the system to adjust the incorrect side. As before, the system automatically recalculates the height and location of the new model.

The next example, shown in 3.15, illustrates a similar procedure to adjust the height of the building. The user needs only to select a point along the base of the building.
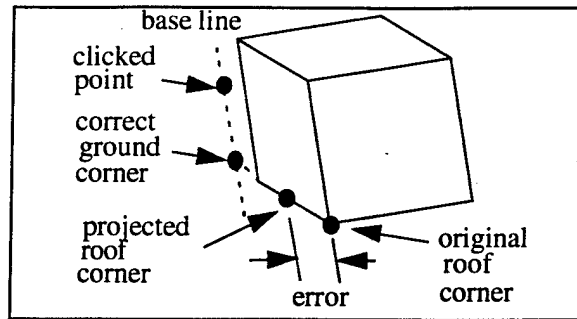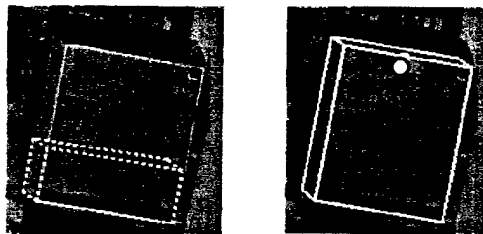
31

Figure 3.13 Illustration of height correction

Figure 3.14 Adjusting wrong side of a building
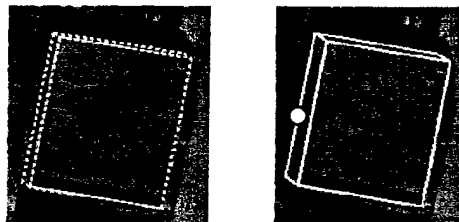

Figure 3.15 Adjusting wrong height of a building

3.3 More Results

This system has been applied to a number of images. Some results from portions of the Ft. Hood, TX site are shown in Figures 3.16 and 3.17 (the Ft. Hood site has been used by several researchers for automatic modeling in recent years, hence provides a good comparison point).

Table 8 summarizes statistical information on the number of clicks and time needed to construct these models. For each roof type (flat or gable), the table shows the number of components that were detected using a given number of clicks. A total of 45 components were modeled, of which only 2 required subsequent height correction.

The total elapsed wall time for this example was 290 seconds. This time does *not* include the initial set up times, such as for selecting an appropriate area to process) or for computing the features that the system uses for automatic analysis. Rather, only the actual time for modeling is included. We believe that these are
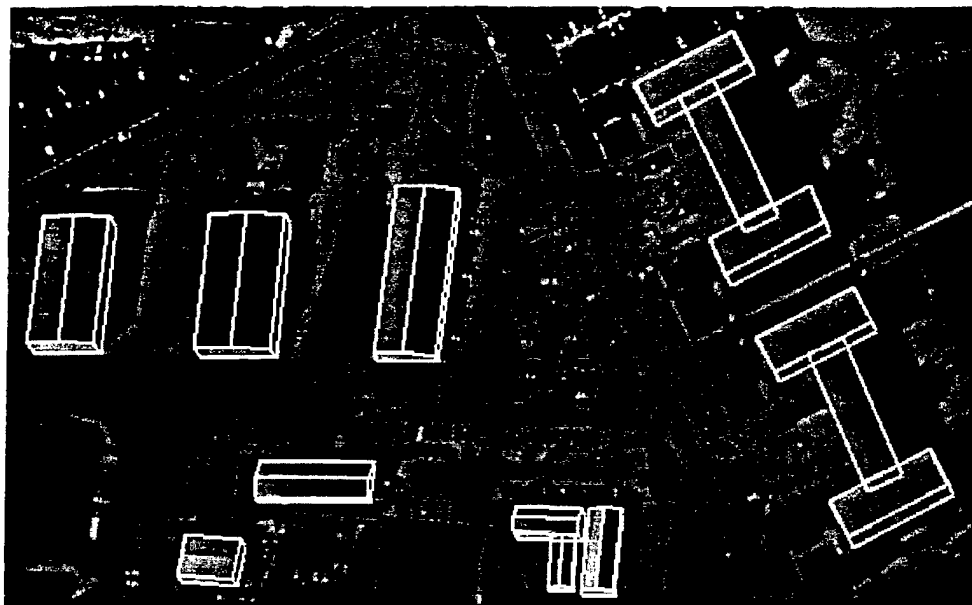
32

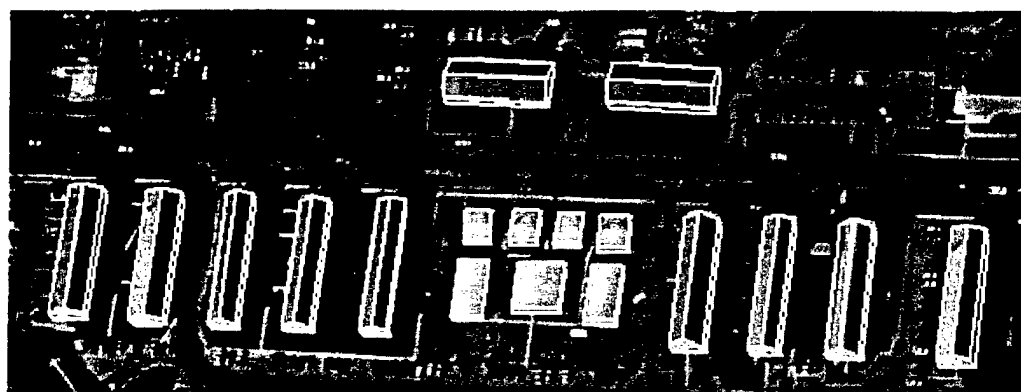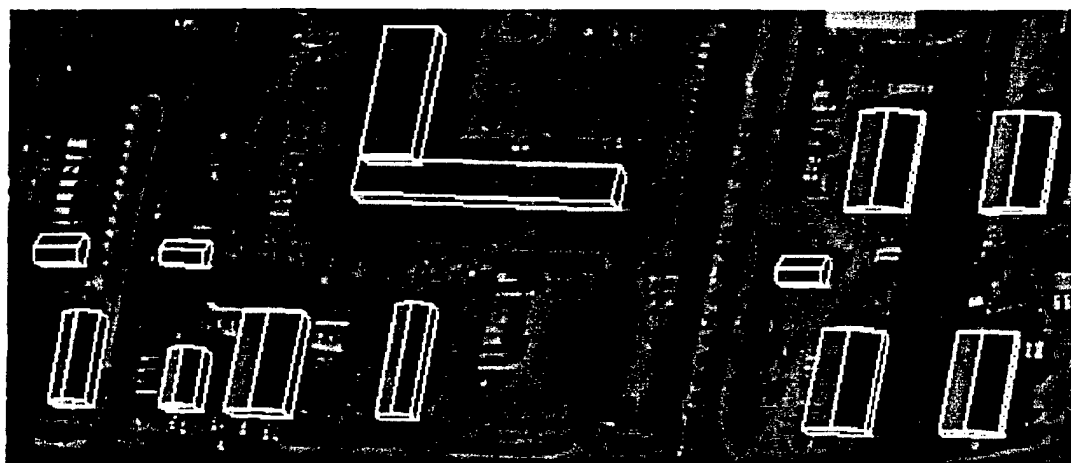**Figure 3.16 Results from Area 1 of the Ft. Hood site**

Figure 3.17 Results from Areas 2 and 3 in the Ft. Hood site

meaningful parameters assuming that a large number of buildings are to be modeled. Our experiments with other images and other sites show similar time requirements.

Table 8: Distribution of Interactions for Three Areas from the Ft. Hood Site

Roof Type	Clicks Needed	Components Formed
Flat	1	1
Roof	2	1
Buildings	3	14
Gabled	2	5
Roof	3	8
Buildings	4	16
TOTAL		45
Buildings requiring height correction = 2		
Total elapsed wall time = 290 seconds		

The results show that the described system can be a very effective tool for constructing models for the classes of buildings within its capability. In our experience, it takes approximately one minute to model each building by using parameter fitting techniques such as provided in [Strat et al., 1992]. In [Grün and Dan 1997], authors report that approximately 30 seconds are required to collect the points needed for recovering a roof description (thought the roof types are more complex than for our system). We have extended our methodology to more complex shapes while still keeping the number of user interactions low; this is discussed next.

3.4 Modeling Complex Buildings

Complex buildings have added dimensions of complexity that include multiple wings, possibly having different heights, and/or arbitrary shapes. Depending on the view point, building sides exhibit a large variety of textures, indentations, and protrusions. This section describes methods to deal with the added complexities that have been recently incorporated into our user-assisted modeling system.

The new methods primarily incorporate facilities to model structures having arbitrary shapes. The outlines of the structures are modeled by polygons that are the result of adding "blocks" to partially constructed models. As with the methods for simpler structures described above, the goal is to minimize the number of user interactions. The new polygonal methods satisfy this goal as the number of pointer (mouse) "clicks" needed is less than the number of corners on the outlines of the buildings.

User interaction starts with a "seed" block. A seed corresponds to a portion of a building that has been either constructed automatically or initiated manually by the user. The user can add or subtract blocks to the seed as needed. The added (or subtracted blocks) can be rectangular or triangular to allow polygonal shapes. These blocks can be specified by one or two clicks.

The system also includes methods to calculate the heights of the various wings when present. Examples are given below and the details of the processes needed to model multicomponent buildings, non-rectangular roof buildings, and multiple height-layered buildings are discussed.

3.4.1 Multicomponent Buildings

The automatic system is designed to model rectangular buildings; therefore, it requires that sufficient evidence be found belonging to the roof. The verification step assumes that evidence for all four sides be present to be counted as positive evidence. Multicomponent buildings have several wings, the roofs of

34

which consists of partial rectangles (Figure 3.18a). While it is possible to add rectangular components to any partial model (Figure 3.18 b, c, and d), the lack of features ("missing" sides and corners in Figure 3.18a) would require that the user issue three clicks per component. This operation would still leave the modeled components separated, requiring a consolidation step to describe the complete structure.
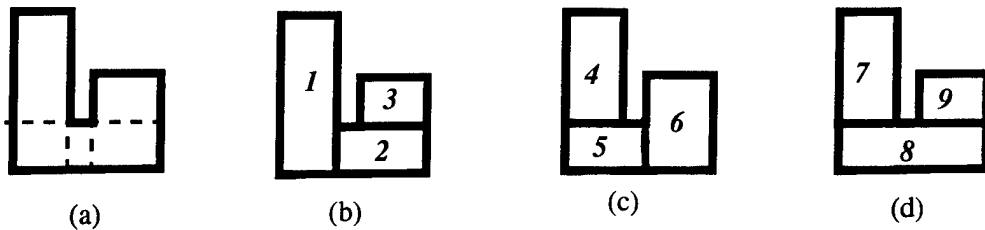

Figure 3.18 Multiwing outline (a) and three possible sets of rectangular components

Instead, methods have been implemented to add or subtract blocks to an existing model. The model "seed" start with a rectangular component (such as any of the numbered rectangles in Figure 3.18) generated automatically by the system, or manually by the user. This process is illustrated in Figures 3.19 and 3.20, for protrusions and indentations, respectively.

Figure 3.19 illustrates the user interaction to add a rectangular block B to block A. One of the clicks is given on the existing outline on block A and the other at the other end of the diagonal across block B. As a result a new polygon C is computed. Figure 3.20 illustrates a similar process for indentations. These operations can be applied repeatedly until the resulting polygon correctly models the building roof. At every step, the system makes use of previously and newly computed evidence to determine the height of the building.
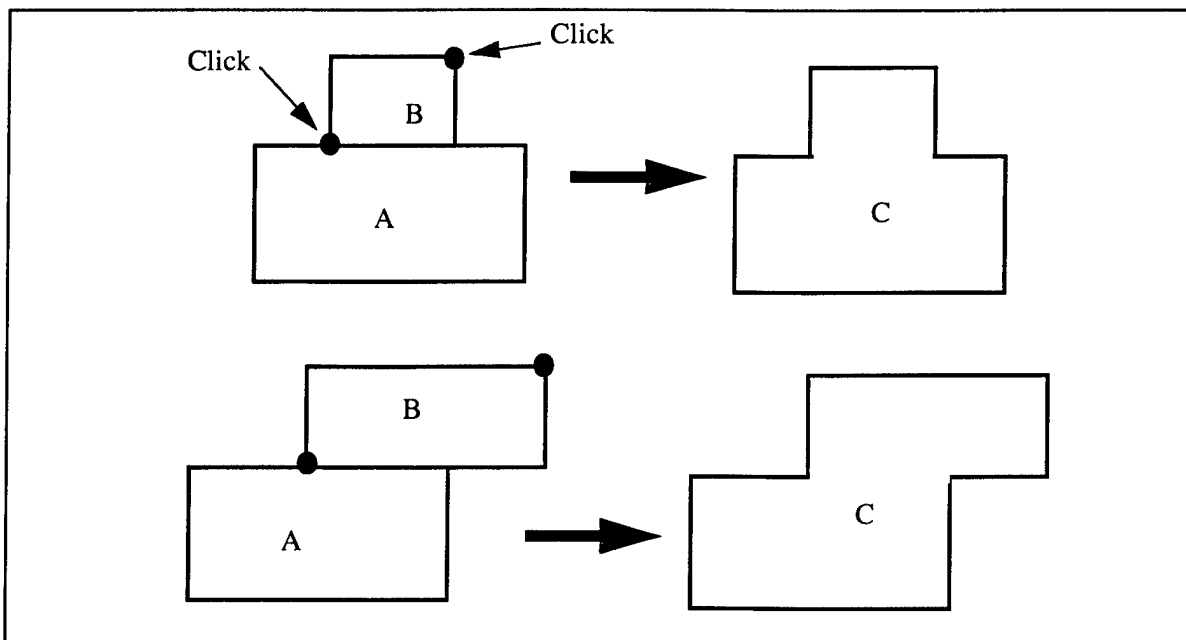

Figure 3.19 Protrusions are added by two clicks

The procedures to generate a protrusion or an indentation are depicted in Figure 3.21 and Figure 3.22 where:
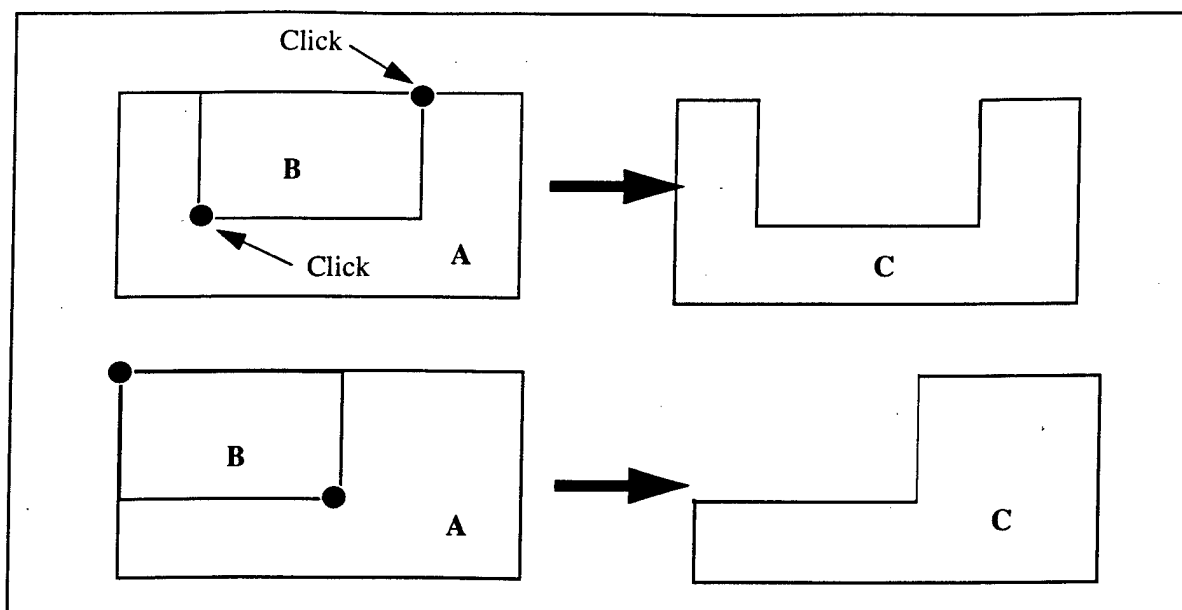
Figure 3.20 Indentations are subtracted by two clicks

L_1: A roof boundary line passing through P_1

L_2: A roof boundary line intersecting L_1

NL_1: A line parallel to L_2, passing through P_1

NL_2: A line parallel to L_1, passing through P_2

NL_3: A line parallel to L_2, passing through P_2

P_3: Intersection point between NL_1 and NL_2

P_4: Intersection point between L_1 and NL_3

This procedure requires 3 clicks for the "seed" block plus 2 times the number of added (protrusion) of removed (indentation) blocks. The height of the modeled structure is made to correspond to that of the seed block. The resulting model can be duplicated and placed wherever similar structures remain to be modeled.

3.4.2 Non-Rectangular Buildings

The methods described above apply to buildings whose corner elements are at, or close, to 90° in 3-D. In order to construct models having non-rectangular sections the system uses triangular blocks. To allow for non-rectangular corners, the angle constraints are relaxed to include junctions between 20° and 160°. To add or remove a triangular section we consider two cases. In the first case only one click is needed to add or remove a triangular protrusion. This is illustrated in Figure 3.23a. The two branches of a junction found near the location of the user's click intersect the boundary of the "seed" model. The second case arises when no such intersection exists, and a second click is required, as illustrated in Figure 3.23.
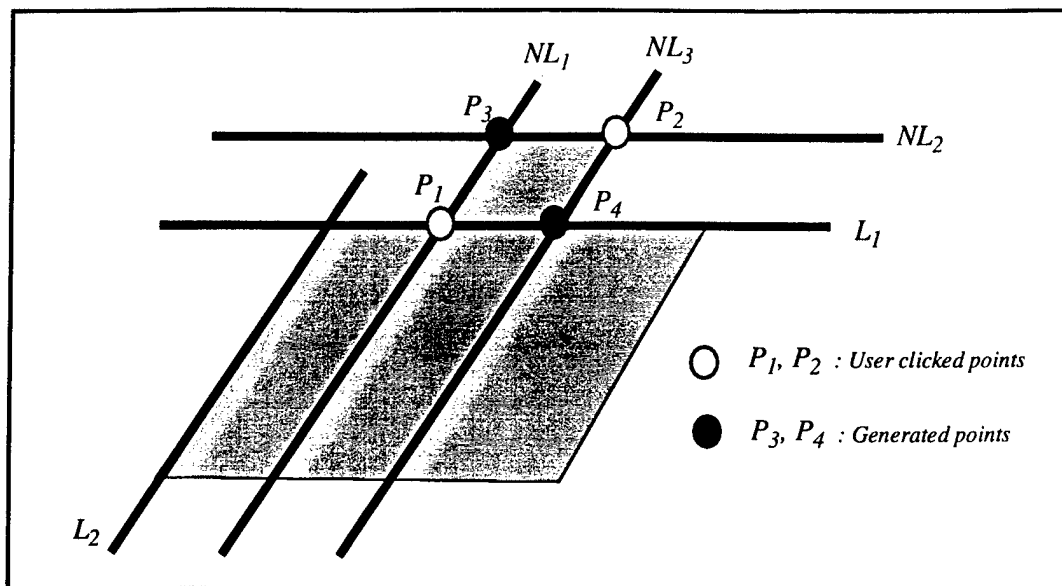
36

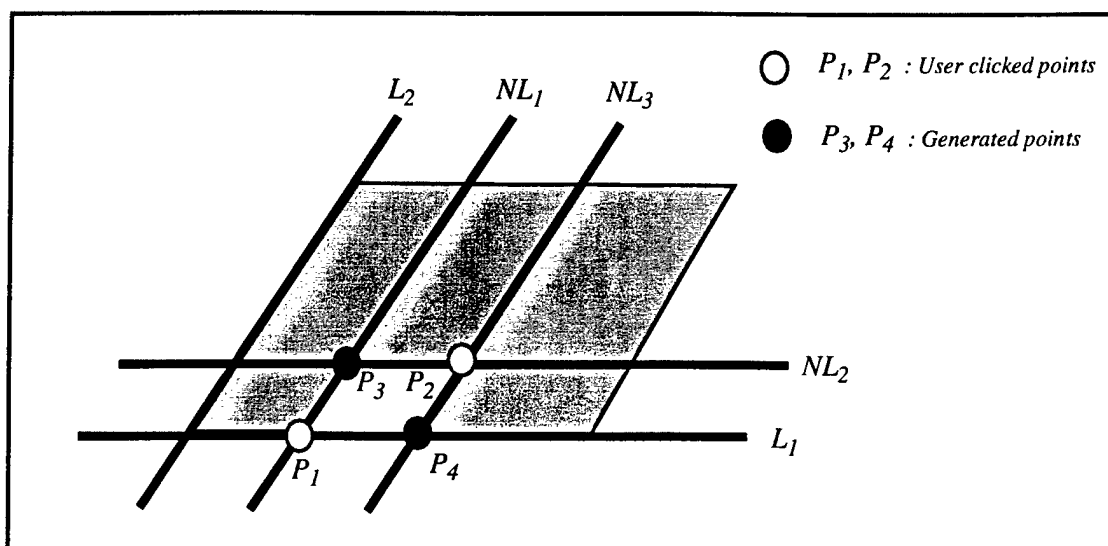**Figure 3.21 Generation of protruding parallelogram given two corners**

Figure 3.22 Generation of indentation parallelogram with two given points

This process can be applied repeatedly to generate the desired 2-D roof boundary. The height of the final model is taken to be that of the initial seed. As before, the seed can be selected from an automatically constructed portion of the model, or generated by the user. Figure 3.24 shows an example of a building in Washington D.C. having indentations and a structure on top. The seed component is shown in Figure 3.24a. For some applications this description would suffice. Other applications may require more detail. After eight clicks of user interaction the four indentations are added, as shown in Figure 3.24b.

A similar result is shown in Figure 3.25, with blocks added to model the shape.
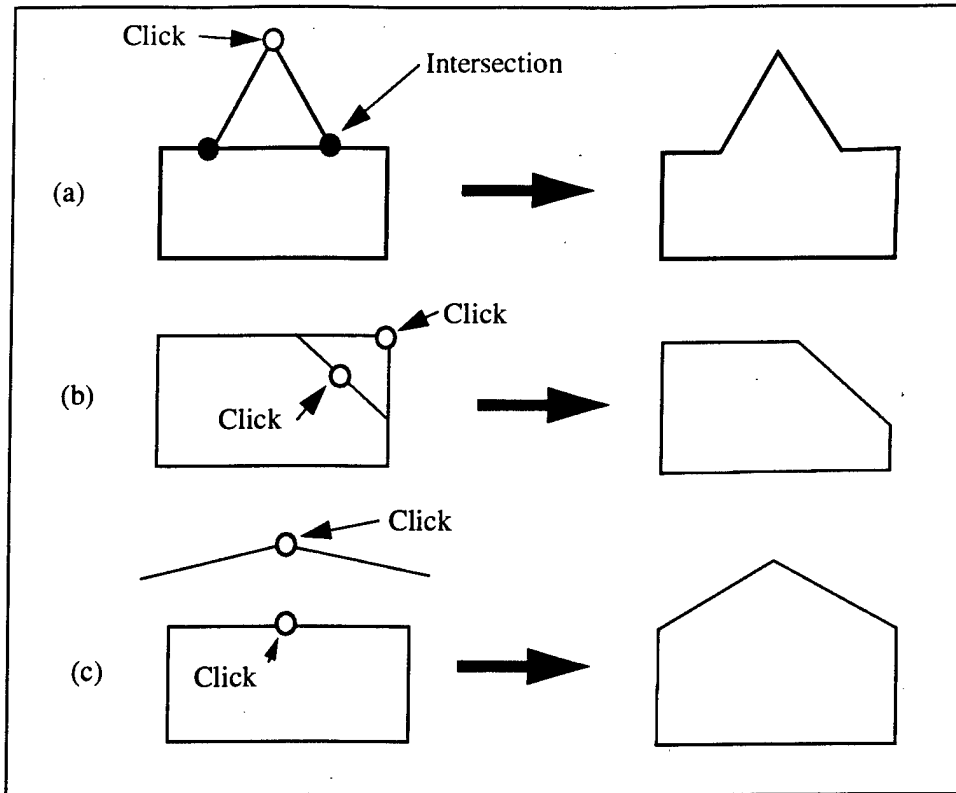
37

Figure 3.23 Adding/removing a triangular block. (a) branches intersect seed. (b) branches aligned. (c) branches do not intersect

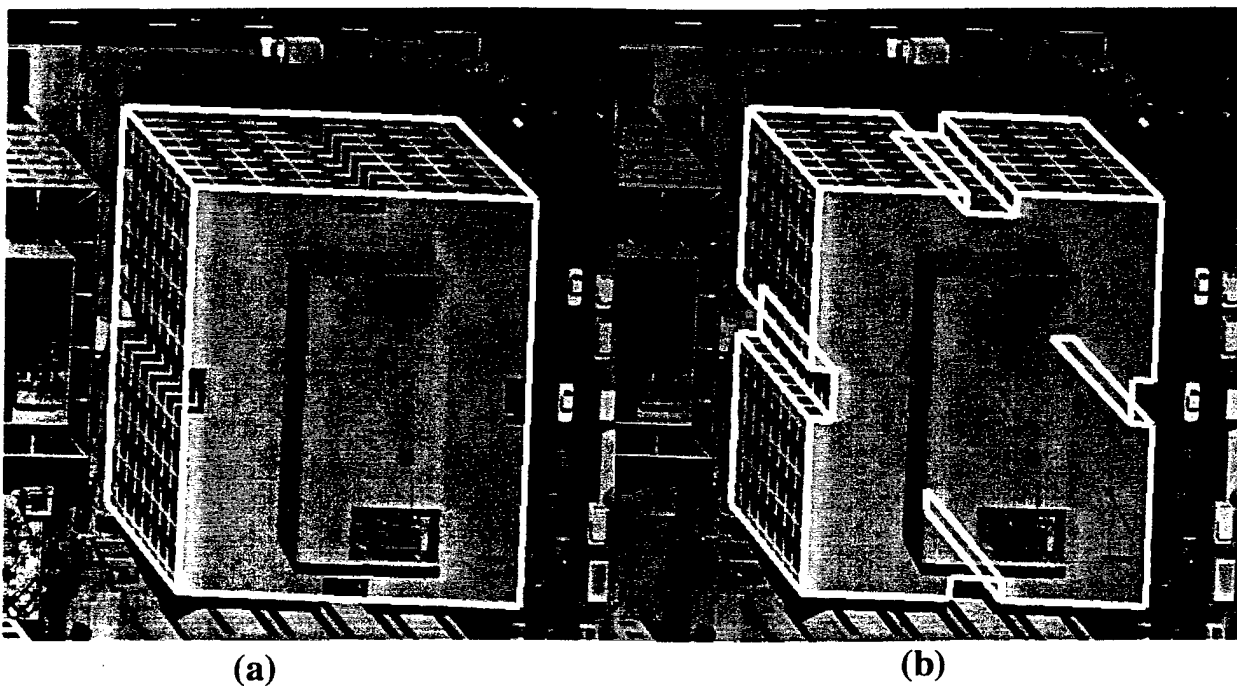**(a)**   **(b)**

Figure 3.24 (a) Seed model (3 clicks). (b) Four indentations (8 clicks) subtracted
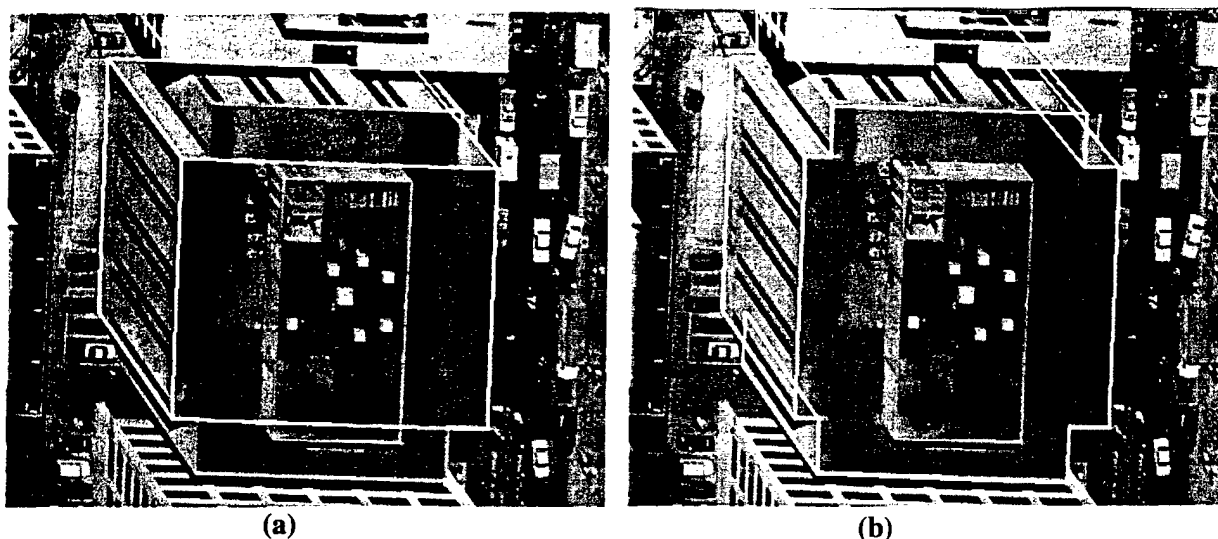
38

Figure 3.25 (a) Seed model (3 clicks); (b) Two added rectangular blocks (4 clicks)

3.4.3 Multilevel Buildings

Complex buildings can have multiple layers, wings at different heights, and superstructures of significant size. In some cases it becomes important to model these as well. Typically the automatic system will detect these separately (or independently) of the main structure. The user-assisted system provides methods to compose these and to adjust their height. Height is considered with respect to the supporting surface. In Figure 3.24, the top layer on the building is found with one click of user assistance. Its height, automatically computed with respect to the ground level, is adjusted automatically as it intersects an existing structure. The final model result for the structures shown in Figures 3.24 and 3.25 is shown in Figure 3.26.

3.4.4 More Results

We have tested the new methods in the user-assisted mode of our automatic building detection and description system using images from the Washington D.C. Mall area. The graphic overlays in the previous and following examples show some of the "hidden" lines in the models.

Figure 3.27 shows intermediate results on a rectangular building. The initial seed takes three clicks (Figure 3.27a) and is augmented by two protrusions requiring four clicks (Figure 3.27b). The top layer requires five clicks (Figure 3.27c).

Figure 3.28 shows intermediate steps in modeling a building having an irregular shape. Figure 3.28a shows the model after the initial seed (three clicks) has been augmented (two clicks) by a rectangular block to form an "L". A triangular block is added (two clicks) in Figure 3.28b. A triangular block is subtracted (two clicks) in Figure 3.28c. A similar procedure constructs the top layer (five clicks) to yield the final model in Figure 3.28d.

Another example of a building of irregular shape is shown in Figure 3.29. In this case the building is partially occluded by and adjacent building and its own top structure occludes part of the main layer. The initial seed (three clicks) plus a rectangular protrusion (two clicks) are shown in Figure 3.29a. The addition of a triangular block (two clicks) in Figure 3.29b is not sufficient to model the pointed protrusion due to its ir-
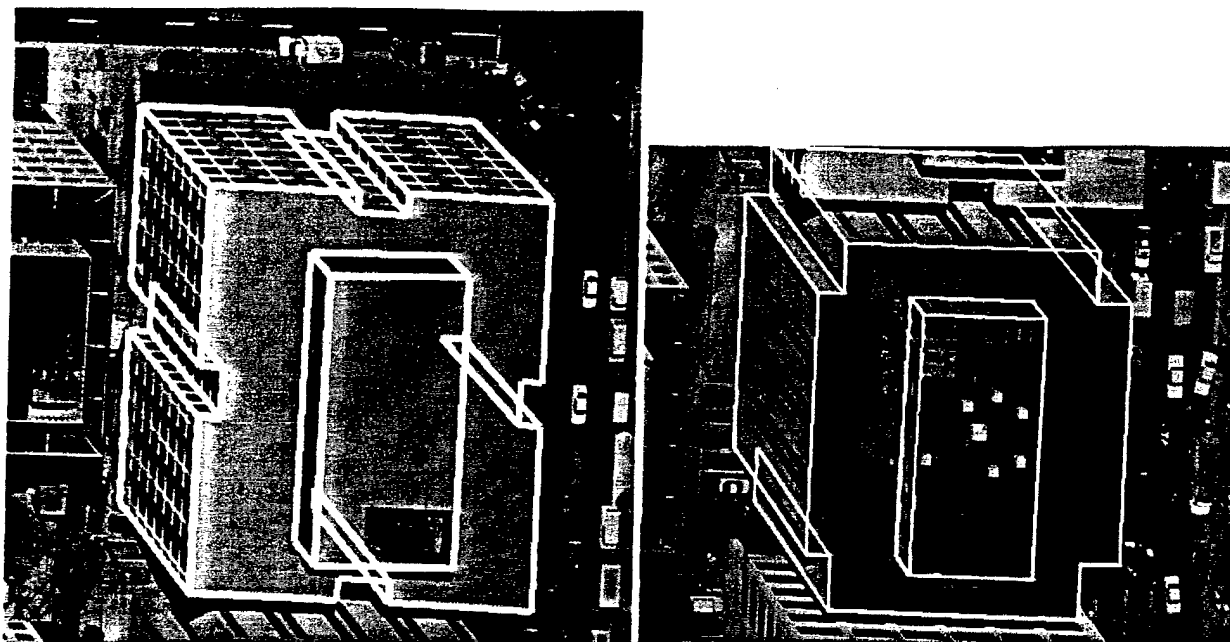
Figure 3.26 Completed models after incorporating top layer (3 clicks each)
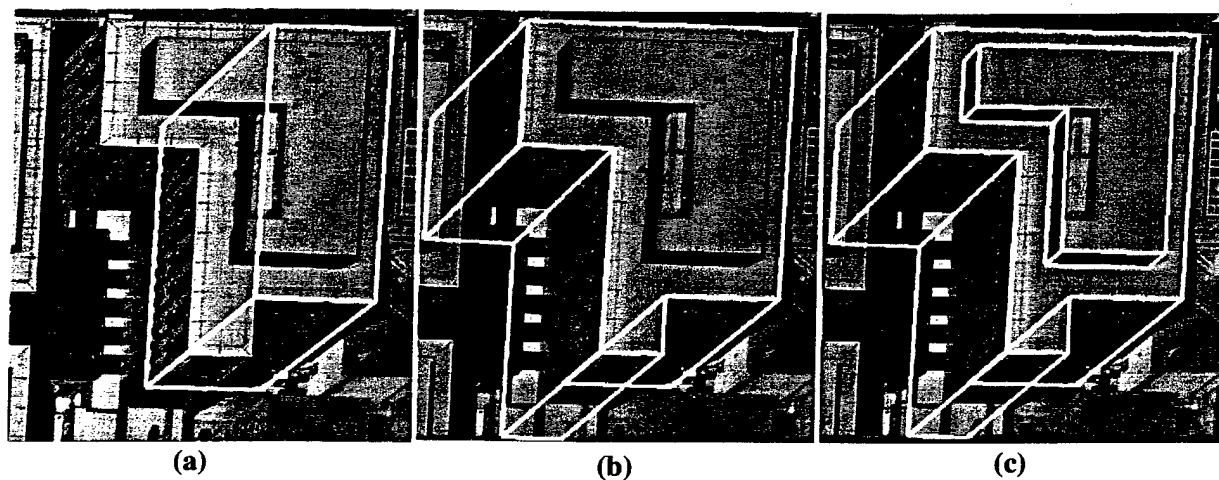

| (a) | (b) | (c) |

Figure 3.27 (a) Seed (3 clicks); (b) Main layer (4 clicks); (c) Top layer (5 clicks)

regular angles. An additional triangular block is added (two clicks) to conform to the shape shown in Figure 3.29c. The completed model with its top layer added (three clicks) is shown in Figure 3.29d.

The last example is shown in Figure 3.30. A cluster of three buildings having irregular shapes and occluding each other. The required user interaction is summarized in Table 9.

40
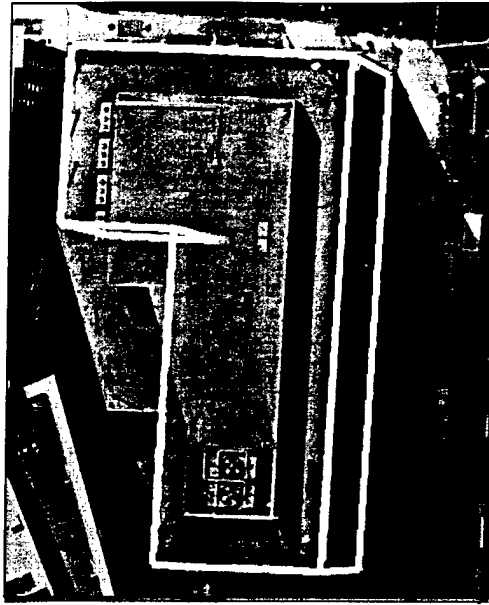
(a) Seed plus rectangular block
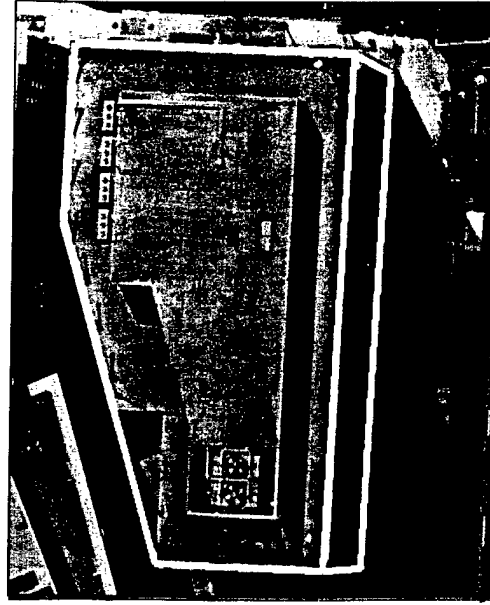
(b) A triangular block added
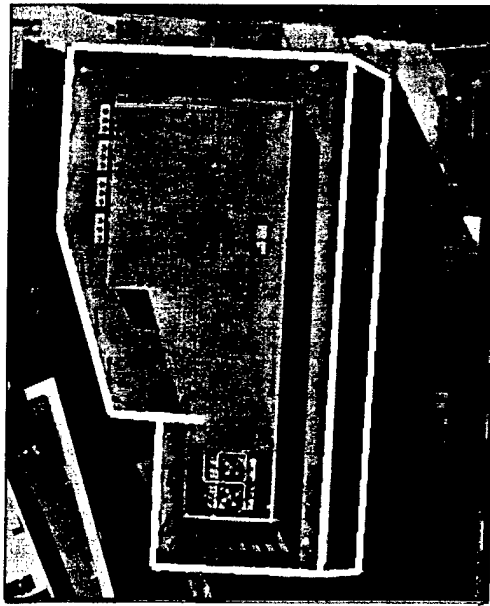
(c) A triangular block subtracted

(d) Top layer added

Figure 3.28 Arbitrary shape. (a) 5 clicks, (b) 2 clicks, (c) 2 clicks, (d) 6 clicks

Figure 3.29 Modeling an irregular shape. (a) Seed plus rectangular block. (b) Added triangular block. (c) Added triangular block. (d) Added top layer. Total: 12 clicks
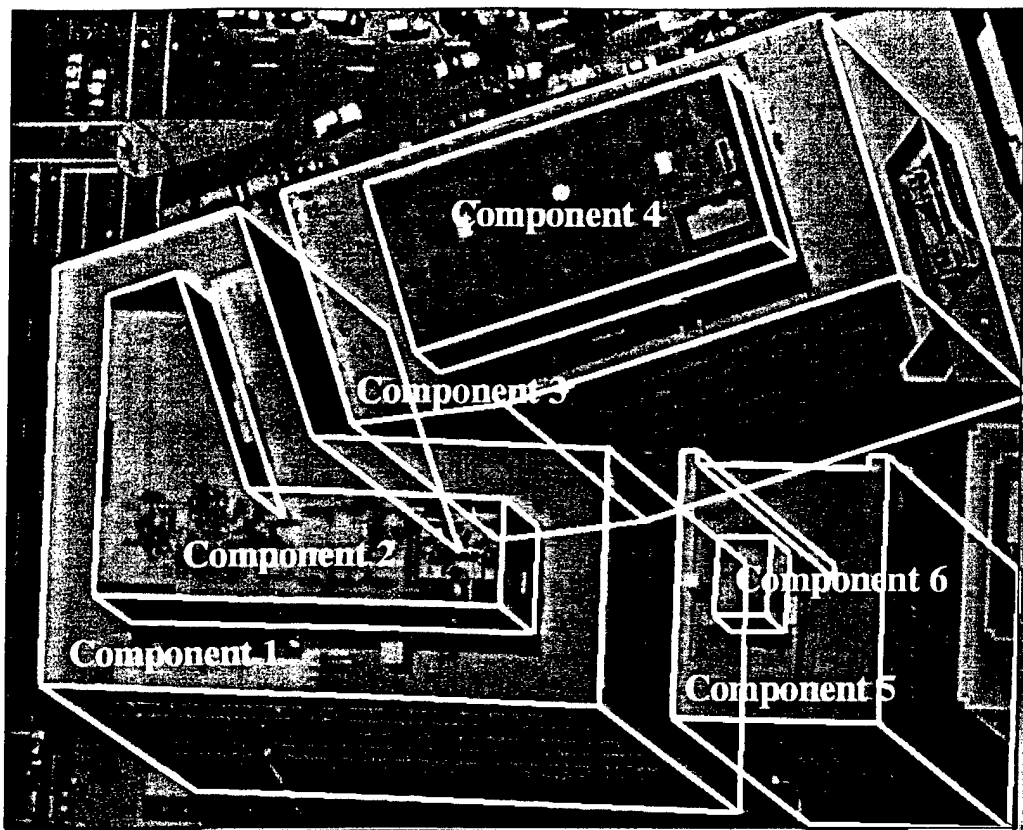
Figure 3.30 Building cluster. To generate this model 28 clicks are required in 30 seconds
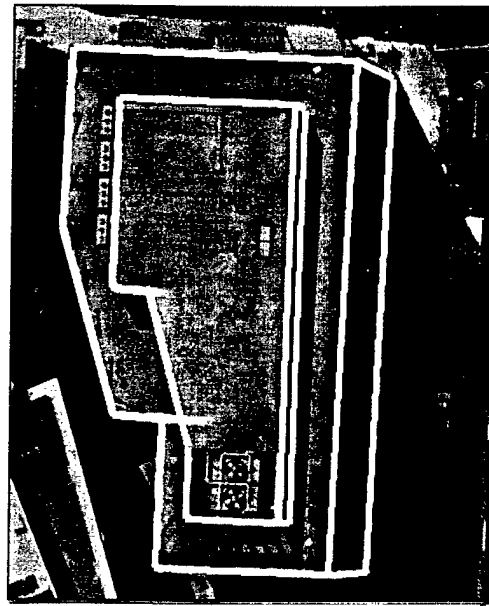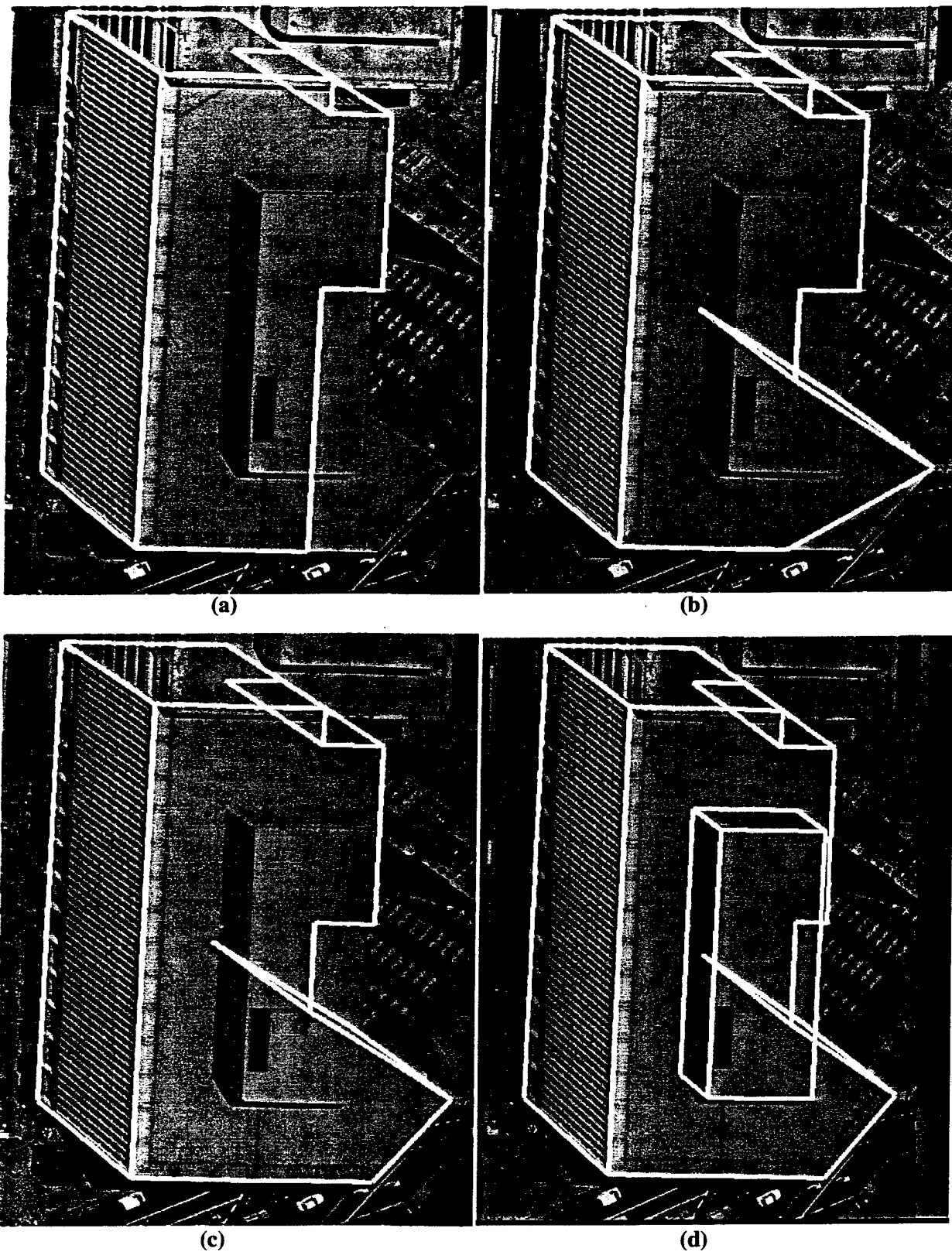
Table 9: User interactions in example of Figure 3.30

Component No.	Clicks	Task	Time (sec.)	Total Time (sec.)
1	3	Seed model	4.5	
	1	Subtract block	0.2	4.9
	1	Add block	0.2	
2	3	Add top structure and adjust height	5.8	
	1	Subtract block	0.2	6.2
	1	Add block	0.2	
3	3	Seed model	5.8	6.8
	2	Subtract block	1.0	
4	3	Add top structure and adjust height	3.6	3.6
5	3	Seed model	5.0	6.2
	2	Subtract block	1.2	
6	3	Add top structure and adjust height	2.1	2.9
	2	Adjust height	0.9	
Total	28			30.6

4. References

[Chellapa et al., 1997] R. Chellapa, Q. Zheng, S. Kuttikkad, C. Shekhar, and P. Burlina, "Site Model Construction for the Exploitation of E-O and SAR Images," in *RADIUS97*, O. Frischein and T. Strat, Ed., Morgan Kaufmann, San Francisco, CA, pp. 185-208.

[Chen et al., 1987] J. Chen, A. Huertas, and G. Medioni, "Fast Convolution with Laplacian-of-Gaussian Masks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*", Vol. PAMI-9, No. 4., July 1987, pp. 585-590.

[Collins et al., 1998] R. Collins, C. Jaynes, Y. Cheng, X. Wang, F. Stolle, A. Hanson, and E. Riseman, "The ASCENDER System: Automatic Site Modeling from Multiple Aerial Images," *Computer Vision and Image Understanding Journal*, special issue on Building Detection and Reconstruction. R. Nevatia and A. Gruen, Ed., 1998, pp. 143-162.

[Curlander & McDonough, 1991] J. Curlander and R. McDonough, "Synthetic Aperture Radar," Wiley Interscience, New York, NY, 1991.

[Fischler et al., 1998] M. Fischler, B. Bolles, and A. Heller. "APGD Evaluation Metrics, Methodology, Rationale," SRI International, May 1998, http://www.ai.sri.com/~apgd

[Haala & Brenner, 1997] N. Haala and C. Brenner, "Interpretation of Urban Surface Models using 2-D Building Information," in *Automatic Extraction of Man-Made-Objects from Aerial and Space Images (II)*, A. Gruen, E, Baltsavias, and O. Henricsson, Ed., Brinkauser, Basel, Switzerland, pp. 213-222.

[Hoepfner et al., 1997] K. Hoepfner, C. Jaynes, E. Riseman, A. Hanson and H. Schultz, "Site Modeling using IFSAR and Electro-Optical Images", *Proceedings of the DARPA Image Understanding Workshop*, New Orleans, LA, May 1997, pp. 983-988.

[Heuel and Nevatia, 1996] S. Heuel and R. Nevatia, "Including Interaction in an Automated Modeling System," *Proceedings of Image Understanding Workshop*, Palm Springs, CA, February 1996, pp. 429-434.

[Hsieh, 1996] Y. Hsieh, "SiteCity: A Semi-Automated Site Modeling System," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 1996, pp. 499-506.

[Huertas et al., 1998] A. Huertas, Z Kim, and R. Nevatia, "Use of Cues from Range Data for Building Modeling," *Proceedings of the DARPA Image Understanding Workshop*, Monterey, CA, Nov. 1998, pp. 577-582.

[Jakowatz et al., 1996] C. Jakowatz, D. Wahl, P. Eichel, D. Ghiglia and P. Thompson, "Spot-Light Mode Synthetic Aperture Radar: A Signal Processing Approach," Kluwer Academic, Boston, MA, 1996.

[Jaynes et al., 1997] C. Jaynes, M. Marengoni, A. Hanson, E. Riseman, and H. Schultz, "Knowledge Directed Reconstruction from Multiple Aerial Images," *Proceedings of the DARPA Image Understanding Workshop*, New Orleans, LA, May 1997, pp. 971-976.

[Li et al., 1998] J. Li, S. Noronha, and R. Nevatia, "User Assisted Modeling of Buildings," *Proceedings of the DARPA Image Understanding Workshop*, Monterey, CA, Nov. 1998, pp. 571-576.

[Maloof et al., 1998] M. Maloof, P. Langlay, and R. Nevatia, "Generalizing over Aspect and Location for Rooftop Location," *IEEE Workshop on Applications of Computer Vision*. Princeton, NJ. October 1998.

[McKeown et al., 1997] D. M. McKeown, S. D. Cochran, S. J. Gifford, W. A. Harvey, J. C. McGlone, M. F. Polis, S. J. Ford, and J. A. Shufelt, "Research in the Automated Analysis of Remotely Sensed Imagery: 1995:1996," *Proceedings DARPA Image Understanding Workshop*, New Orleans, LA, May 1997, pp. 779-812.

[Nevatia, 1999] R. Nevatia, "On Evaluation of 3-D Geospatial Modeling Systems," in ISPRS *Proceedings of the International Workshop on 3D Geospatial Data Production*, Paris, France, April 1999.

[Nevatia & Huertas, 1998] R. Nevatia and A. Huertas "Knowledge-Based Building Detection and Description: 1997-1998," *Proceedings DARPA Image Understanding Workshop*, Monterey, CA, Nov. 1988, pp. 469-478.

[Noronha & Nevatia, 1997] S. Noronha and R. Nevatia. "Detection and Description of Buildings from Multiple Aerial Images," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, PR, June 1997, pp. 588-594.

[Roux & McKeown, 1994] M. Roux and D. McKeown, "Feature Matching for Building Extraction from Multiple Views," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, June 1994, pp. 46-53.

[Stetina et al., 1994] F. Stetina, J. Hill, and T. Kunz, "The Development of a Lidar Instrument for Precise Topgraphic Mapping," *Proceedings of International Geoscience and Remote Sensing Symposium*, Pasadena, CA, August 1994.

[Strat et al., 1992] T. Strat, L. Quam, J. Mundy, R. Welty, W. Bremner, M. Horwedel, D. Hackett, and A. Hoogs, "The RADIUS Common Development Environment," *Proceedings of the DARPA Image Understanding Workshop*, San Diego, CA, 1992, pp. 215-226.